REPORT RESUMES

ED 014 332 24 PS 000 240

ACCELERATION OF INTELLECTUAL DEVELOPMENT IN EARLY CHILDHOOD.
FINAL REPORT.
BY- BEREITER, CARL
ILLINOIS UNIV., URBANA
REPORT NUMBER CRP-2129 PUB DATE JUN 67
REPORT NUMBER BR-5-0385
CONTRACT OEC-4-10-008
EDRS PRICE MF-$1.00 HC-$8.48 210P.

DESCRIPTORS- *INTELLECTUAL DEVELOPMENT, PRESCHOOL LEARNING, *PRESCHOOL CHILDREN, KINDERGARTEN CHILDREN, *PRESCHOOL EDUCATION, CONSERVATION EDUCATION, TEACHING MACHINES, VERBAL LEARNING, DISADVANTAGED YOUTH, READING INSTRUCTION, PROBLEM SOLVING, COGNITIVE DEVELOPMENT, *PROGRAM DEVELOPMENT, ITPA, MONTESSORI, START,

THE CHILD'S CAPACITY FOR SELF-ACTUATED INTELLECTUAL GROWTH AND THE POSSIBILITY OF SPEEDING UP INTELLECTUAL GROWTH THROUGH IMPROVED OPPORTUNITIES AND INCREASED STIMULATION WERE STUDIED. SIX EXPLORATORY STUDIES CARRIED OUT DURING THE FIRST TWO YEARS OF THIS PROJECT WERE REPORTED. THE THREE MAIN AREAS OF LEARNING WHICH WERE INVESTIGATED WITH THE IDEA OF LOCATING PROMISING APPROACHES WERE READING, CREATIVITY, AND LOGICAL OPERATIONS. THESE STUDIES CONCERNED (1) EXPLORING THE TEACHING OF READING TO VERY YOUNG CHILDREN, (2) A TEACHING MACHINE APPROACH WHICH SHOWED SOME PROMISE IN THE FIRST STUDY, (3) PREFERENCES FOR HIGH-FREQUENCY VERSUS LOW-FREQUENCY WORD USE OCCURRING IN CHILDREN'S SPEECH, (4) CONSTRUCTION ACTIVITIES INVOLVING INDEPENDENT PROBLEM-SOLVING, AND GUIDED CONSTRUCTION, (5) A METHOD OF INDUCING CONSERVATION OF SUBSTANCE IN KINDERGARTEN CHILDREN, AND (6) TEACHING FORMAL LOGICAL OPERATION TO PRESCHOOL CHILDREN. TWO OTHER STUDIES WERE DISCUSSED, INCLUDING (1) INSTRUCTION OF DIRECT VERBAL INSTRUCTION IN LANGUAGE, ARITHMETIC, AND READING TO FOUR-YEAR OLD DISADVANTAGED CHILDREN, AND (2) COMPARISON OF A DIRECT VERBAL INSTRUCTION WITH A MONTESSORI PROGRAM FOR FOUR-YEAR OLDS. RESULTS AND CONCLUSIONS WERE MANY AND VARIED. (EF)

ED014332

BR 5-0385

PA 24

FINAL REPORT

Project No. 2129

Contract No. OE 4-10-008

ACCELERATION OF INTELLECTUAL DEVELOPMENT IN EARLY CHILDHOOD

June 1967

U.S. DEPARTMENT OF

HEALTH, EDUCATION, AND WELFARE

Office of Education

Bureau of Research

PS 000240

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ACCELERATION OF INTELLECTUAL DEVELOPMENT IN EARLY CHILDHOOD

Project No. 2129
Contract No. OE 4-10-008

Carl Bereiter

June 1967

The research reported herein was performed pursuant to a contract with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

University of Illinois
Urbana, Illinois

PS 000240

Preface

In their journals and professional meetings early childhood educators devote a great deal of time to reassuring themselves and others that they truly love children. I have noticed that even scientists, when addressing this audience, will grasp some opportunity to let it be known that deep down inside, and scientific objectivity notwithstanding, they too love children. Since some such testimony seems to be a requirement for being taken seriously, I hereby render mine: In all my adult years I have encountered only about ten people I didn't like, and no more than two or three of them have been children. I believe that this proportion is respectably low, considering that children do not generally go out of their way to make themselves agreeable or to hide their faults.

The important discovery, which first made itself widely known during the eighteenth century, that children are human just like the rest of us, became exaggerated during the Romantic era into the belief that children are more human than the rest of us. This belief still has a powerful hold on the popular imagination, and no doubt accounts for the sanctimonious insistence on a reverential attitude toward children. Behind this belief, however, I detect in many devoted early childhood educators an abhorrence for the modern adult world, with its "discipline of the machine," its timetables, its relentless purposefulness, and its insistence on logic. Such educators turn in dread from any suggestion that early education should be a calculated preparation for the rigors of adult life.

If an attitudinal test is to be required of people who would like to have a part in the education of young children, I would like to see it contain some indication of how a person feels about adulthood as well as childhood.

To establish the precedent, I hereby offer a second testimonial: The improvement in the quality of existence that can take place between early childhood and maturity seems to me as great as the improvement that must take place phylogenetically between the level of the rat and the level of the human child. A popular and, as usual, groundless belief has it that half of adult intelligence is acquired by age four (Leonard, 1966). But the child who dies at age four has not lived half a life. He has scarcely begun the great adventure, available only to homo sapiens, of establishing mastery through his intelligence over a world that dwarfs him in every other respect. With each generation the magnitude of this adventure grows, both in its risks and in its potential triumphs.

Few of us are as fit as we might have been for this adventure. Most people have to settle for being spectators, some find the adventure so distressing that they will have nothing to do with it at all, and some of these, unfortunately, seek refuge in the world of the child, where the child's ignorance becomes their delight.

The new curriculum movements that have exploded in the last decade seem to have received their impulse, not from new facts, but from a sudden shift in the valuation of what was already known to be the case. For a long time educators had been concerned with the large number of pupils who failed in high school mathematics. But they were finally made to realize (largely through college teachers of mathematics) that the pupils they had counted as successful were also failures, that what had passed for knowledge and intelligence was, by another standard, ignorance and stupidity. On consideration of what the children were being taught in high school mathematics, it became evident that there was no reason why this sorry state of affairs had to continue. This shift in valuation has since spread over the whole range of academic content and levels. The normatively average student, who had so long enjoyed the blessing of educators (providing he was also emotionally and socially well-adjusted), was thenceforth judged to be illiterate, unthinking, uninformed, and out-of-date.

Sputnik has been blamed for this shift in valuation, but it could only have triggered what had to come. Very simply, the modern age in its totality requires that anyone who would participate in it as an active member be more intelligent than people have been in the past. This requirement is naturally a bit perplexing to the educator who thinks of intelligence as something partly delivered through the genes, and otherwise a matter of how much love and attention one has received as a young child. There is no evident way of drastically altering the gene pool or upgrading the love and attention children receive. But what is meant by upgrading intelligence in general is the same as what is meant by upgrading mathematics ability in particular: it means teaching children more of what is most valuable, and teaching it better. With older children it is not so difficult to see what this means, although it may be difficult to bring it about. But with preschool children it is not at all evident how this notion applies. We are used to thinking of them as developing rather than learning, as being nurtured rather than being taught. For older children logic is a subject; in younger children it is a trait.

If there is a discontinuity between school-age and preschool children, however, it stands to reason that it must be a discontinuity in our thinking rather than in the natural order. If we find it difficult to think of early childhood education in terms of rate of learning, structure of disciplines, definable content, prerequisite concepts and skills, clarity of statement, and so on, it is perhaps only because people before us have not bothered to try. It is only as we learn to do so that we shall be able to talk purposefully about improving early education to make children more intelligent.

My hope is that the studies reported here will make some contribution to eliminating the widening gap between our understanding of curriculum and instructional problems in education at higher levels and our understanding of these problems at the preschool level, where the question, "What did you learn at school today?" is so incongruous that only a child could fail to appreciate its absurdity.

The eight studies making up this report were carried out over the course of three years; what with the coming and going of personnel between and during studies, some 40 people had a direct hand in their execution. Some of the people who had major parts are identified as joint or sole authors of sections of the report--Siegfried Engelmann, David Brison, Martin Bender, and the late Charles Summers. Two others whose contributions to the project have been great in value and extent are not so identified, however--Philip Reidford, who was significantly involved in most of the major undertakings of the first two years, and Jean Osborn, who joined the project at the time the first full-scale preschool was conceived, and without whom it could never have developed into what it has. Others who can be singled out for special contributions at various points along the way are Mary Lou O'Connor, Elaine C. Bruner, Roger Gehlbach, Kay Case, and Ingeborg Rose. It remains to find some way of indicating, within the brief span of a sentence, the profound and many-faceted contributions of Siegfried Engelmann. To retreat into the indefinite, I shall say that if anything in this report strikes the reader as meritorious, there is a good chance that most of the credit is due to him; I will go further and say that he is even entitled to a share in its faults.

Carl Bereiter
Urbana, Illinois
February, 1967

Contents

Preface

1.0 Introduction

- 1.1 Nature of the study
- 1.2 Historical background

2.0 Exploratory studies

- 2.1 Teaching reading to two- and three-year-olds
  - 2.11 Problem and background
  - 2.12 Objectives
  - 2.13 Procedures
  - 2.14 Results and discussion
  - 2.15 Conclusions
  - 2.16 Summary
- 2.2 The effect of free-time use of a START teaching machine on reading ability in the kindergarten (co-author, Martin Bender)
  - 2.21 Problem
  - 2.22 Design and procedures
  - 2.23 Results
  - 2.24 Discussion and conclusion
  - 2.25 Summary
- 2.3 Children's preferences for high versus low-frequency words (co-author, Charles Summers)
  - 2.31 Problem and objectives
  - 2.32 Procedure
  - 2.33 Results
  - 2.34 Discussion and conclusions
  - 2.35 Summary
- 2.4 Four approaches to construction activities in the nursery school and their relation to creative problem solving
  - 2.41 Problem
  - 2.42 Objectives
  - 2.43 Method and procedure

        2.44 Results
        2.45 Discussion
        2.46 Summary
    2.5 Teaching conservation of substance (by David W. Brison)
        2.51 Prefatory note (by Carl Bereiter)
        2.52 Problem and background
        2.53 Method
        2.54 Results
        2.55 Discussion and conclusions
        2.56 Summary
    2.6 Teaching formal operations (by Siegfried Engelmann)
        2.61 Prefatory note (by Carl Bereiter)
        2.62 Nature of the study
        2.63 Method
        2.64 Results
        2.65 Discussion
        2.66 Summary
    2.7 Resume of exploratory studies
3.0 Studies in direct verbal instruction
    3.1 An academically-oriented preschool for disadvantaged children (co-author, Siegfried Engelmann)
        3.11 Introduction
        3.12 Method and procedures
        3.13 Results
            3.131 Summary of test results
        3.14 Discussion
        3.15 Conclusions
        3.16 Summary

3.2 Direct verbal instruction contrasted with Montessori methods in the teaching of normal four-year-old children
    3.21 Purpose
    3.22 Method and procedure
    3.23 Results
    3.24 Discussion
    3.25 Summary
3.3 Implications for future research in early education

1.0 Introduction

1.1 Nature of the study

This study started out to be psychological and ended as a study in pedagogy. In retrospect, it is possible to discover a foreshadowing of this mutation in the original statement of problem, which is here quoted in full from the project application:

> The importance of the early childhood years for personality and social development has long been recognized by psychologists, and important changes in child-rearing practices have resulted. The possible importance of these years for intellectual development has until recently received much less attention. This disparity of interest is reflected in our nursery schools which, however much they may contribute to the socialization of children, touch only incidentally on matters relating to intellectual development.
>
> A recent upsurge of interest in early cognitive development appears to have its intellectual roots in (1) a belated recognition in this country of the significance of Piaget's work, with its emphasis upon the active and complex intellectual life of the young child, (2) the evidence from a variety of educational experiments, but particularly those of O. K. Moore with the teaching of reading to very young children, that young children are capable of much higher intellectual attainments than had previously been thought and that these can be achieved without coercion or strain on the child, simply by capitalizing on his exploratory and problem-solving propensities, and (3) a number of human and animal studies showing profound and far-reaching effects of experiential enrichment and deprivation.
>
> It appears that we have scarcely begun to explore the potentialities for acceleration and enhancement of early intellectual development. What is needed at this time is a great deal of what may be called "innovative research"--that is, research which tries out a variety of new ideas and remains open and flexible enough to pursue any leads which seem promising. This is a proposal for a series of six studies of this nature. These studies should contribute to our presently scanty knowledge of techniques for accelerating cognitive development in young children.

The problem is couched in the language of developmental psychology. The key issues suggested in paragraph two are the child's capacity for self-actuated intellectual growth and the possibility of speeding up intellectual growth by improved opportunities and increased stimulation. The final paragraph, however, suggests that there may be more to it than this, that speeding up intellectual growth (whatever that might mean) requires more than simply throwing stimulating material in the child's way, that some technology is needed. Although it is certainly not clear from the above paragraphs, I think I was conscious at the time that children become smarter through learning things, and that therefore "techniques of accelerating cognitive development" had to involve ways of teaching young children things they didn't already know and that would be of some use to them in subsequent learning. The implications of this view lead straight to pedagogy.

However uncertain the rationale, it is the case that, instead of setting out to stimulate intellectual growth in some general fashion, we attempted from the start to teach children various concepts and skills that we thought might be important. In the course of these attempts, the first of which were quite fumbling, we quickly lost interest in such generalities as cognitive development, enrichment, stimulation, and the like. It became evident that reading, for instance, whether it is presented at age three or at age seven, is still the same task. There are certain things that have to be learned regardless, and once one begins to consider how to teach these particulars to the children one is dealing with, such notions as level of development, enrichment, and stimulation lose their relevance. This is not such a radical notion in regard to reading; but if it is true of reading, why is it not also true of anything else a child may learn? May it not be the case that in teaching children to reason, to be creative, and so on, one is also dealing with tasks that have certain requirements regardless of the learner, and that the problem is always a pedagogical one of devising curricula and teaching methods that will enable the child to meet these requirements? This is the view we were led to in the course of this investigation; as a result the original roots in developmental psychology were severed and new roots were established in the subject matter of learning.

Section 2 of this report spans the transition from psychology to pedagogy. It includes three of the six studies originally proposed: a pilot study in the teaching of reading to very young children, which was concerned with presentational modes, but which was conceived with hardly a thought to the particular things children of any age would have to learn in order to read; a study in creative problem-solving applied to construction activities, again concerned with "treatment variables" rather than with content; a study in teaching conservation of substance--here, however, unsuccessful pilot work led eventually to a consideration of what children had to learn in order to be able to handle the task, and thence to a successful approach by David W. Brison, in which conservation was treated as a learning task involving identifiable problems of motivation, attention, and comprehension like any other. Whereas the original proposal had called for further work with concrete operations, Siegfried Engelmann jumped ahead to teach a formal logical algorithm to preschool children, and in so doing developed a prototype for the teaching of advanced subject matter to young children through the careful use of analogic structure and verbal rules.

The second phase of investigation, reported in Section 3, took off from this point. Having completed the transition to a pedagogical approach to intellectual development, the tempting course of action was to abandon the question of whether children could learn this or that and try to get some idea of how much children could learn from a fast-paced, content-centered curriculum. Disadvantaged children were chosen for the first study, as ones least likely to be learning the material out of school. Middle-class children were used in a second study, to provide comparative information, and also to take advantage of the opportunity of measuring the effects of the program against those of a different approach, more closely wedded to a mentalistic view of intelligence--the Montessori method.

1.2 Historical background

If one chose to take a "great man" view of educational history, then early childhood education could be regarded as the lengthened shadow of Jean Jacques Rousseau. In Rosseau's Emile, written over two hundred years ago, we find stated with amazing clarity the ideas which even today form the creed of the "modern" educator in his battle against the forces of traditionalism. The following quotations are taken from Boyd's (1962) translation:

> People sometimes speak about a complete man. Let us think rather of of a complete child. (p. 65)
>
> The wisest people are so much concerned with what grown ups should know that they never consider what children are capable of learning, They keep looking for the man in the child, not thinking of what he is before he becomes a man. (pp. 5-6)
>
> Begin then by studying your pupils better; for assuredly you do not know them. (p. 6)
>
> This is nature's way. Why set yourself against it? Do you not see that in attempting to improve on her work you are destroying it and defeating the provision she has made? (p. 17)
>
> The only habit the child should be allowed to acquire is to contract none. (p. 22)
>
> May I set forth at this point the most important and the most useful rule in all education? It is not to save time but to waste it. (p. 41)
>
> ...the first education should be purely negative. (p. 41) So far as you can, refrain from a good instruction for fear of giving a bad one. (p. 43)
>
> ...before you dare undertake the making of a man you must be a man yourself. (p. 42)
>
> No Verbal Learning (p. 46) For what in fact do they [the professional pedagogues] teach? Nothing but words. (p. 47)

> Reading is the greatest plague of childhood. (p. 51) Our first masters of philosophy are our feet, our hands and our eyes. To put books in place of this experience does not teach us to reason: only to be credulous and to borrow the reason of others. (p. 54)
>
> He learns his lessons from nature and not from men. (p. 54)
>
> Make your pupil attend to the phenomena of nature, be in no hurry arouse his curiosity. But to nourish this curiosity, be in no hurry to satisfy it. Suggest problems but leave the solving of them to him. Whatever he knows, he should know not because you have told him, but because he has grasped it himself. Do not teach him science: let him discover it. (p. 73)

Rousseau's ideas involved a number of psychological, anthropological, and sociological speculations. They antedated by a century the scientific study of these areas of knowledge and they were not in any sense prophetic of subsequent findings. Their survival, therefore, stands as testimony to the lack of significance that these sciences have had for educational theory.[1]

Piaget has replaced Roussau's simple notions about the emergence of mental faculties with an elaborate theory of genetic epistemology, but when it comes to drawing inferences about what this means for educational practice, the conclusions emanating from the Institute Jean Jacques Rousseau do not differ appreciably from those of its namesake (Piaget, 1964). Even among those psychologists who will credit him with little else, Freud is supposed to have administered the coup de grace to the eighteenth century notion, shared by Rousseau, of the rational man. Yet the prescriptions for child-rearing set down by one of the modern psychoanalytic thinkers who has gone farthest in translating ego psychology into operational terms read, in large part, as if they were lifted word-for-word from the Emile (Dreikurs, undated).

---

[1] A recent article in Look Magazine (Leonard, 1966), illustrates the perennial rediscovery of these ideas. Virtually every one of the Rousseaunian ideas cited above is put forth in this article as the latest thing on the educational horizone,the fruit of recent psychological discoveries.

Rousseau did not offer a technology of education, a systematic set of rules that could be applied without additional invention to bring about desired results.

> My plan is not to go into details, but only to set forth general principles and gives (sic) examples in difficult cases. (p. 44)

It remained for others to devise techniques. The quaint sensory training devices of Maria Montessorri (1912) and the electronic age implements of O. K. Moore (1963) are alike in their Rousseauian conception. They are not techniques for controlling or shaping the child but for controlling and shaping the environment in which the child autonomously learns. The efforts to build modern science and mathematics curricula around "discovery learning" are in the same vein. Whereas Rousseau had to take Emile to the same hilltop for many sunsets and sunrises to point out to him a phenomenon that might pique his curiosity, the modern science teacher can produce a device such as a pendulum on a spinning platform that will dramatize an interesting phenomenon more handily, but the principle of "Do not teach him science: let him discover it" remains the same. The role of the teacher in all of this is not to teach but to arrange encounters between the child and the physical world from which learning will emerge. Rousseau objected to trying to teach children what they would need to know in the future on the grounds that the child would probably not live to make use of it (p. 33); today's educators object on the grounds that the child will probably outlive the usefulness of the knowledge we have to impart. In practice it comes to the same thing.

To Rousseau's eighteenth-century mind, systematic pedagogy was an example of the irrational attempt to improve upon nature, whereas the sane thing to do was seek accord with nature. The industrial revolution of the following century took place over the dead bodies of defenders of this view, although the view survives today among what C. P. Snow (1963) has called the "new Luddites." Technology had been developing since prehistoric times, but it may be said that in the nineteenth century people discovered what technology was about and began to develop a technology for technological progress itself. The secret, as Whitehead has noted, was to start with the desired end result and work backward through a series of steps that were capable of bringing it about, rather than starting with things as they were (the "natural order") and looking for some way to alter them. Thus, the steam locomotive and the automobile were not ways to improve upon the horse, but ways to "harness" the most efficient sources of power in the service of locomotion. Nature was something to be used rather than improved upon. Paradoxically, this burgeoning technology was applied to virtually every area of human activity except that which dealt with human behavior itself (Skinner, 1958). Thus it was possible for pedagogy to remain, as it remains for most people today, embedded in a pre-nineteenth century way of thinking. It is this, I think, which explains the amazing currency of Rousseau's ideas and their immunity to the advance of scientific knowledge.

The incongruity of this survival of eighteenth century thought in mid-twentieth century America is dramatically illustrated in Project Head Start. The War on Poverty, of which Head Start was to be a major instrument, represents the typical twentieth century faith in technology. A previously unrealized goal, the abolition of poverty, was set, and means were sought to bring it about. This was quite a different matter from the philanthropic efforts of earlier generations merely to improve upon things as they were. Early education of children of the poor was seen as a necessary step toward the attainment of one of the requirements of a povertyless nation--the breaking of the tie between the intelligence of children and the education of their parents. But what could an eighteenth century conception of education offer? An appeal to "nature's way," to negative teaching, to the avoidance of verbal learning, to a rejection of concern for "what grown ups should know" in favor of a concern over "what children are capable of learning." The children of the poor, therefore, were to be educated, in so far as the intellect was concerned, in the same manner as Rousseau's Emile, the fictional son of a wealthy nobleman, whose only income-producing skill was to be that of a carpenter! It is ironic that Rousseau drew much of his inspiration from the peasantry, whose character and style of life he tended to idealize. Thus twentieth century educators, when called upon to assist in eliminating the remnants of peasantry and its urban counterpart, were drawing upon an educational philosophy which glorified those very traits that made people peasants, that barred them from advancement in a society built upon technological sophistication and verbal intelligence.

For those who recognized that the problem might require different measures for its solution, the course taken to find new measures was precisely that prescribed by Rousseau: "Begin then by studying your pupils better; for assuredly you do not know them." The pre-industrial belief, again, that the way to improve transportation was by studying the horse. Innovations in education were attempted, but, true to Rousseauian premises, they consisted of altering the environment, enriching its content of objects and devices. The child was still to go his own way.

Educational research today may be characterized as an attempt to leap from the eighteenth century to the twentieth century without passing through the nineteenth. Its achievements to date would suggest that this is not possible. By far the largest part of educational research has been devoted to "studying our pupils better," to describing with increasingly sophisticated techniques their course of development and to identifying the distinguishing characteristics of the underachiever, the non-reader, the gifted child, and so on. This remains the most active area of educational research today. The second most active area, to judge, for instance, by papers submitted for the annual American Educational Research Association convention, is that which deals with characteristics of the school environment--classroom atmosphere, teacher characteristics ("before you dare undertake the making of a man you must be a man yourself"), and the frequency of various sorts of teacher acts. The research techniques are modern, but the paradigms of the educational process within which they are applied antedate the technological era.

On the practical side, education gives every indication of entering, experimentally at least, the electronic age. Television sets appear in school classrooms, but the face on the screen is that of Emile's tutor. The teaching machine gives an even stronger impression of twentieth century technology. Beginning as a typical nineteenth century apparatus of gears and levers, it has matured into a computer-controlled system making use of a variety of advanced hardware. Still, available evidence seems to indicate that it is the content of the instruction purveyed by machine or text that makes the difference rather than characteristics of the medium. The critical aspect of this new educational technology is the breaking down of some large unit of learning, such as "fractions" or "Russian grammar" into discrete constituents each of which can be learned in their order and which are sufficiently well defined that it is possible to test whether they have been learned before instruction proceeds to the next item. As Cronbach (1963) has observed, the main contribution of the machine or "programmed text" medium is that it forces the program writer to perform this analysis rather than approaching instruction in the haphazard way that ordinary textbook writing or curriculum planning allows. But for this kind of analysis of instructional units, no technology exists. The program writer must proceed as intuitively as he would have had to centuries ago.

It is the program and curriculum writers, engaged in the laborious task of analyzing learning units into their necessary constituents, who may be said to be working their way through the nineteenth century. It does not matter that they are trying to teach twentieth century science or mathematics. The same job could have been done a century ago with the science and mathematics of that day. The point is that it was not done then, and it has to be done before anything that can be called a technology of instruction will be realized.

For what the program writer does is set a learning goal and then devise a series of steps leading up to it. It is the same kind of working backward from end through successive means that took place in the projects large and small which added up to the industrial revolution. That the industrial revolution had to do with machines is only a result of the fact that its concerns were largely with the processing and transportation of goods. Its essence was the way in which previously unrealized goals could be pursued through efficient means. Thinkers of the past had found it difficult to conceive of ways of pursuing previously unrealized goals except through appeal to analogy or through utopianism. Appeal to analogy often failed because it was difficult to find analogies that were close enough--consider the failure to achieve human flight through the application of analogy with birds. Utopianism consisted of envisioning a set of conditions that would make attainment of the goal possible, but without being able to indicate how the conditions themselves might be achieved. Rousseau's scheme of education relied heavily on both analogy and utopianism. He saw the infant as analogous to an animal and the child from two to twelve as analogous to the savage, and he drew many of his proposals from what he believed to be true of the development of animals and savages. Two conditions essential for the realization of his educational ideal were the isolation of the child from societal influences and the possession of extraordinary qualities by the tutor, both of which Rousseau himself recognized as impossible of attainment (pp. 19, 42).

It would be inaccurate to suggest that no technological progress in education has been made during recent centuries, just as it would be inaccurate to say that no industrial technology had developed before the nineteenth century. Between the Middle Ages and the middle of the twentieth century, the teaching of reading had progressed from a kind of black art to a set of regular procedures that appear to be successful in about 80 per cent of cases. Such technical advances as have occurred, however, have been more in spite of than through the agency of the kind of analogical and utopian thinking engaged in by Rousseau. With considerable scorn, Rousseau noted,

> Great stress is laid on finding better methods of teaching children to read. Reading cases and cards have been invented, and the child's room has been turned into a printer's shop. Locke suggested the use of dice. Fancy all this elaborate contrivance! A surer way that nobody thinks of is to create the desire to read. (pp. 51-52)

A worthwhile suggestion, perhaps; but Rousseau apparently did not consider the possibility of developing a technology for creating the desire to read.

The most ambitious attempt to develop a technology of instruction was that of the Jesuit educators, embodied in the Ratio Studiorum, which was issued in 1599 and revised from time to time over the next 200 years. True to the spirit of technology, the authors of the Ratio Studiorum sought neither to obey nature nor to improve upon it, but rather to use it for their purposes. For instance, rivalry among students (the program was designed for older pupils) was deliberately exploited by assigning each pupil a competitor or by dividing the class into two armies who competed for scholastic honors. The approach to improving educational methods was empirical rather than inventive, however. The idea was to search out the best of what was already being done rather than to set new goals and devise ways of reaching them (Curtis & Boultwood, 1956, pp. 151-158).

It seems fair to say that most technical progress in education since the Ratio Studiorum has been of the same kind--empirical rather than inventive. In recent years objective research methods have partly taken the place of expert judgment in identifying the best of existing techniques, but they have not led to any greater emphasis on invention. In fact, some curriculum planners have claimed that objective methods of assessment have tended to discourage invention, since they do not allow for the possibility that innovations in method may outrun the development of objective criteria for evaluating them (Atkin, 1963). What is perhaps more crucial, however, is the fact that crude empiricism does not allow for intrinsic differences in the perfectibility of methods. To use our well-worn example of the horse and the motor car, the first motor cars would probably not have fared very well in empirical comparisons with the horse. Certainly this was true on the criterion of reliability. The difference was that the horse was about as good as it could become, whereas it was evident to anyone who understood the design of the motor car that it contained within it the potentiality for far-reaching improvements. In the early stages of technological development, perfectibility is usually a far more important criterion than immediate performance. As Whitehead (1948, p. 91) noted, for invention to flourish, it is important to be able to look ahead to the next step in technological advance. It is this which makes possible what Whitehead (1948, p. 92) has aptly called "disciplined progress," progress which may be pursued professionally without having to wait for "the occasional genius, or the occasional lucky thought."

The notion of an educational technology is easily misunderstood. At the risk of some repetitiousness, I shall try to indicate what a technology of early education would mean and contrast this with the situation that actually exists. A technology of early education does not necessarily entail the use of mechanical or electronic gadgetry, although it may. It does not mean converting an art into an exact science. It has more to do with how art is applied. The first requirement of technology is that remote goals or desiderata must be cast in the form of problems or difficulties that can be dealt with one at a time (Whitehead, 1948, p. 92). Moreover, it must be possible to relate the problems to one another, as well as to the ultimate goal, so that when a problem appears insurmountable, it can be replaced by alternate problems or reduced to sub-problems; also, the problems should be related in such a way that when one problem is solved it is clear what new problems can or should be tackled, making use of the solution that has been achieved. In this way, orderly progress is possible.

The beginnings of a technology of reading instruction are evident in analyses such as that of Carroll (1964). The task of reading, which in the mature reader is executed as if it were a single coherent process of extracting meaning from printed characters, is broken down by Carroll into separate problems and sub-problems that can be treated individually yet without losing sight of the way in which these problems are related to one another. The hierarchical analyses carried out by Gagne and his associates (Gagne, et al., 1962) on mathematics learning tasks are also examples of the beginnings of "orderly progress" in the development of instructional methods. In neither case do the analysts provide much in the way of suggested solutions to the separate learning problems; but the striking thing about such analyses is that the more thorough they are, the more problems and difficulties they are able to isolate, the less formidable seems the task of dealing with these problems and difficulties.

By contrast, early education has hardly reached the stage where problems are recognized at all, let alone the stage where large problems have been broken down into smaller problems that can be dealt with in a straight-forward manner. Ask a nursery school teacher what her problems are and she is likely either to state them in particularistic terms--such as what to do about Jimmy Smith, who kicks people--or in terms of ideals--such as the fostering of curiosity, readiness for reading, conceptual development, a positive self-image, mental health, or cooperativeness. These are not problems in the sense of which technology treats of problems. They are, rather, concerns which at best provide the motivation to search for problems, as yet unidentified. It is possible to deal with these concerns, and sometimes to good effect; but the successful dealing with them does not lead to technological advance. The teacher who succeeds in dealing with Jimmy Smith, who kicks, may in the process acquire some personal savvy that permits her to deal more successfully at a later time with Ronald Jones, who bites; but the skills so acquired either die with her or are passed on to other teachers through a master-apprentice relationship. A policy or general approach may develop that is successful in fostering curiosity in children; but unless there is some way of determining what problems have been solved and how, what problems remain unsolved, what new problems have been created, and what possibilities for new attainments are opened by the solutions that have been achieved, disciplined and steady progress is impossible. Further progress remains accidental and identifiable only after it has happened.

The inability to meet the challenge of Project Head Start in a straightforward and purposeful manner is thus easily understood. Starting with the larger project of the War on Poverty, we can see the technological mind at work.. The problem of abolishing poverty was broken down into a number of sub-problems, of which one was rendering children of the poor capable of succeeding in school and thus reaping the economic benefits that accrue to the well-educated person. This is the kind of broadly stated mission that the non-specialist may be expected to hand down to the specialist, on the assumption that the specialist will possess the technological sophistication necessary to convert it into a program of problem-solving efforts that will result eventually in a way of accomplishing the mission. This is what we see happening, for instance, as scientists and engineers work toward the mission of getting a man to the moon and back. After several years of work, no man has as yet gotten anywhere near the moon, nor even tried. What is happening, however, is that difficulties standing in the way of accomplishment of the mission are being identified and overcome, and new problems are not only being encountered but are being sought out systematically, so that the whole effort has a clear direction to it, even though no one could have set down, in the early stages of the project at least, a complete and adequate recipe for accomplishing the mission.

Superficially, it might appear that the same kind of thing has been going on with respect to the mission of enabling poor children to succeed in school. A large amount of research has been going on, which has turned up a good deal of information about the children in question. A variety of educational programs are being tried and their effects being measured. But if we ask the same kinds of questions that can be asked of the moon mission--what difficulties are known to stand in the way of its accomplishment, which ones have been overcome and which have not, what the problems are in overcoming the remaining difficulties, what efforts are being made to seek out and anticipate other difficulties, and what the indicated next steps are for bringing the project to fruition, what sort of timetable the project is on, etc.--some revealing differences come to light. In the first place, the mission of compensatory preschool education (to give it a title) was approached as if it could be accomplished straight-away. It was as if the space scientists, being handed the moon mission, had proceeded immediately to launch all the equipment they happened to have available on the outside chance that something would reach the moon, and then had reflected soberly that although none of the efforts was entirely successful there was some merit in each of them. What problems other than administrative have been solved? None. What problems other than administrative are we now closer to solving today than we were in 1964? It is impossible to say. What stumbling blocks other than administrative have we discovered that were not foreseen? What, in fact, are the problems that need to be solved before we can succeed in the mission of enabling poor children to succeed in school?

This last is the most important but also the trickiest question, for it brings into doubt the relevance of all the research that has been devoted to early learning and to disadvantaged children. To be sure, we can say much more about young children in general and poor children in particular than we could a few years back, but does this allow us to state any problems in a way that we can tackle them? We also know much more about the environment beyond the earth's atmosphere than we did a few years back; but this knowledge has proved its worth by allowing us to set aside some problems that people had thought might be serious but have turned out not to be and by enabling us to define other practical problems much more precisely.

It is necessary to distinguish two kinds of problems. One, we may call epistemic problems; they represent obstacles in the path of our knowledge or understanding of phenomena. The other, we may call technical problems; they represent obstacles in the path of our attainment of some practical objective. The two kinds of problems are related but not coterminous. There are epistemic problems concerning conditions in outer space--the distribution of radiation, for instance. But radiation may or may not present technical problems for space travel--problems that have to be contended with in the designing of equipment, the plotting of routes, and so on. When technology is in an advanced state, the relations between epistemic problems and technical problems are clear-cut and tend to be taken for granted. But when technology is primitive the relations are obscure; vast amounts of knowledge about the phenomena one is dealing with have no practical relevance whatever, and it is seldom clear what kinds of knowledge would be of benefit. The research that has been carried out on young and disadvantaged children has been directed toward epistemic problems. We know that lower-class Negro children give preferential status to white dolls, that their scores on most verbal tests are about a year below average, that they live in crowded housing, that they leave final consonants off words, that they are more physically aggressive and ask fewer questions than middle-class white children, that most of them have not been inside museums or travelled on escalators, that they tend not to call each other by name, and that they have usually not been read stories or nursery rhymes, but that they do watch television a lot (see Ausubel, 1964). What we do not know is whether any of these things make any difference _for the teacher_ who must educate them; or if they do, what kind of difference and how much.

It would be getting things backwards to criticize the research that has been done for its lack of practical relevance. Rather, it is the technological backwardness of early education, its inability to formulate problems of a technical nature, that makes it impossible for research to have relevance to it. The problems of compensatory preschool education that can be formulated are those that I have indicated earlier: particularistic problems concerning individual children or unique situations, and mere designations of areas of concern or interest. Given this pre-technological state of the pedagogical art, it seems quite inappropriate to plea for more research to help guide compensatory preschool education. There is no evident way that it could draw guidance from research.

I have concentrated upon early education as it applies to disadvantaged children because here the urgency of public concern dramatizes the inadequacy of our present state of technology. With respect to early education of the normal child, even the concerns are vague--a feeling on the part of some zealous parents that nursery schools and kindergartens should be "doing more" for their children, and a feeling on the part of forward-looking people that the future holds challenges that early education should somehow be mobilizing itself to meet.

The context of technological impoverishment that I have sketched is one in which almost any attempt to set specific goals for early education and to pursue them methodically is liable to appear radical and unnatural. This is true not only of attempts to employ teaching machines and deliberate behavior modification procedures, but also of attempts to apply curriculum and teaching practices that would appear quite ordinary at higher school levels. On the other hand, because of the lack of any cumulative know-how, it is not easy for people to distinguish those departures from custom that represent mere whim or trial-and-error from those that have some rationale or practical sense to them. For that reason, it is common for innovators to dignify their proposals with appeals to psychological theories and precepts, even though the theories and precepts are so remote from practical decisions that they can be used indifferently to support any and all proposals. Thus, in reaction to the preschool program that we have developed through the present project, we have been variously congratulated for having put into practice the ideas of Bruner, Ausubel, Hunt, Piaget, Skinner, and Chomsky; and we have also been specifically criticized for "flying in the face of" the truths set forth by these authorities. There is nothing paradoxical in this: it merely illustrates the point that at the present time in early education there is little correspondence between the procedures and problems involved in producing behavioral phenomena and the procedures and problems involved in explaining them.

2.0 Exploratory Studies

The six studies reported in this section were carried out during the first two years of the project. They were not intended to have continuity or convergence. Instead, they were conceived of as probes carried out to locate promising areas of early learning and promising approaches, to be made the subject of more penetrating investigation later. Three main areas of learning were investigated: reading, creativity, and logical operations.

The first study explored the teaching of reading to very young children. It was small in scope and lacking in experimental control. As the report will show, the original methods were so modified in the course of the study that they lost their distinctive identities, but it was through just these modifications that the only findings of interest arose.

The second study was distantly derived from the first. A teaching machine approach that had shown some promise in the first study was tested in kindergartens under controlled conditions, but it was used only as an optional, free-time activity by the children and thus represented a very minimal kind of educational intervention.

The third study is the only one that could properly be termed basic research, in that it sought to answer a question about children's preferences for certain kinds of words rather than to test the feasibility or effectiveness of an educational approach. As it turned out, our later work in reading did not make use of sight-vocabulary learning, and so the results of this study had no application within the project; but they may be of interest for programs that do emphasize sight vocabulary.

The fourth study was a short-term experimental comparison of four methods of handling construction activities in the nursery school. It was perhaps a premature attempt at rigor in this area, since more might have been gained by a looser study that attempted to develop a single effective means of producing behavioral change.

The fifth study was preceded by some exploratory work, and was a controlled test of a method of inducing conservation of substance in kindergarten children.

The sixth study had a dual purpose. On the one hand it investigated the possibility of teaching a formal logical operation (which children would not normally be expected to master until near adolescence) to preschool children. On the other hand the study involved a comparison between quite privileged and disadvantaged children on a set of tasks that put extraordinary demands on their ability to use language in reasoning. It was this latter, more-or-less incidental aspect of the study that provided a starting point for the second phase of this project.

2.1 Teaching Reading to Two- and Three-Year-Olds

2.11 Problem and background

For several reasons, reading seemed to be the logical area of early learning to focus upon in studying acceleration of intellectual development. Of all areas of school learning it had been the most thoroughly studied from the point of view of the age at which it could be mastered. After learning to talk it was the most significant intellectual attainment of the childhood years. It therefore seemed that the extent to which the age at which children learned to read could be lowered would serve as a marker of the extent to which intellectual development could be speeded up in other areas, and that very likely the approach that was most successful in speeding up the acquisition of reading would be the approach that was most successful generally in promoting early learning.

The mental age of six and a half had, with some research support, come to be widely accepted as the age at which a child could be expected to handle reading instruction successfully (Dolch & Bloomster, 1937). To lower this age by a year or even two through a superior concentration of educational resources would not be a particularly significant demonstration, however. Children of such mental ages had been turning up in first grades for years, and some of them had made satisfactory progress in reading. But to teach reading successfully to children with mental ages below four would indeed indicate remarkable possibilities for speeding up intellectual development.

There were hopeful indications that this could be done. Moore (1963) had been starting children below the age of three at learning to read through a voice and typewriter system that allowed the child to explore the relations between printed and spoken symbols independently. Fowler (1962) had published a doctoral dissertation on teaching a two-year-old child to read; and Doman, Delacato, Williams, and Doman (1963) had published in a woman's magazine an article which claimed to provide a tested method of teaching reading to children, starting as early as nine months.

Of the reports available, however, only Fowler's provided any data, and that only on one subject. Although his subject did begin reading instruction at two years, six months, she was of exceptional intelligence (IQ ranged from 170 to 150 on various testings), and so had a mental age of nearly four at the time instruction began. The child did make substantial progress in acquiring sight vocabulary during the early months of instruction, but after seven months she began to show signs of increasing tension and the instruction was terminated. At the age of four and a half she showed renewed interest in reading, instruction was resumed, and the child soon learned to read. By that time, of course, she had attained the "normal" mental age for learning to read.

Moore's results have consistently been reported in terms of unjoined particulars--some children start below age three, some children are of below average intelligence, some children accomplish such-and-such before the age of six, and so on. Thus it is not clear whether any children in Moore's studies learned to read at remarkably low mental ages.

Finding a feasible way of teaching reading to very young children would be a significant accomplishment, however, regardless of the mental age of the children, because of the procedural and motivational problems involved. Children of two and three cannot be expected to have the definite desire to learn to read that children of school age often exhibit, nor to put up with the kind and amount of work that are called for in regular school reading instruction.

It seemed obvious that reading instruction with very young children would have to be carried on through play-type activities. The approaches that others had taken did, in fact, correspond to use of the two major media of children's play--games and toys. Fowler and Doman et al. had used typical parent-child games as a vehicle for reading instruction. Moore's "Talking Typewriter," was functionally similar to many simpler toys that command the interest of children because of the complexity of the control that the child can gradually attain over it--marionettes, gyroscopes, and magnetized dogs, to name simple instances, and model railroads, to name an example which in its most elaborate versions approaches the complexity of the talking typewriter itself. In attaining control over any of these toys the child has to attain a functional mastery of the rules according to which the toy operates, and so by building a "toy" that operated according to standard orthographic rules, Moore had constructed one such that in attaining control over it the child had to learn rules that also enabled him to read.

Although both the game and the toy approaches are based upon the child's normal play activities, they have considerable pedagogical differences. The parent-child game approach, as represented in Fowler and in Doman _et al._, is essentially a means of keeping the child engaged in activities that will provide him with continued exposure to and some inducement to learn reading materials. It is actually very much like the more relaxed types of reading instruction in the schools, except that there is much more frosting and less cake. It makes up for what the child may lack in goal orientation by abundant social reinforcement and sheer fun. The toy approach, on the other hand, provides little continuing motivation of these kinds and must rely heavily on the child's motives to understand and to control.

This notion of "games" and "toys" as the two fundamental ways of structuring learning for young children provided the rationale for the present study.

2.12 Objectives

This study was intended as a small, exploratory comparison of two approaches to teaching reading to very young children, one built around social games and the other centering upon the use of teaching machines. The purpose was to develop as we went along a workable program of each kind that could be used in a larger and more formal experimental comparison later, while collecting clinical observations on the responses of children to each approach.

2.13 Procedures

Subjects: Although the eventual aim of the research on teaching reading was to test methods on average three-year-old children, practical considerations made it advisable to do the pilot work on less typical but more accessible children--namely, children of university faculty and students. Because such children could be expected to be intellectually advanced for their ages, two-year-olds as well as three-year-olds were recruited. Five children at each age level were obtained. The two-year-olds were assigned to receive instruction by the teaching machine approach, because we were more concerned about the cognitive problems associated with teaching machines and therefore wanted the teaching machine group to be comparable in mental maturity to the target population of average three-year-olds. The three-year-olds were assigned to the social games approach, because with this approach social behavior variables seemed likely to be more important, and thus it was desirable to have the pilot subjects the same chronological age as the target group.

Three weeks after treatment began, the subjects were given the Illinois Test of Psycholinguistic Abilities. Results of this testing are reported in Table 2.13.1 for all subjects. As expected, the subjects did tend to be above average in test performance. Although all of the children were below four years chronologically, six of them obtained a total language age score exceeding four years. Only one child obtained a language age score below his chronological age. As it turned out, there was little difference in the language ages of the two-year-olds and the three-year-olds. The mean language age was 3-10 for the younger group and 4-0 for the older--a difference of only two months, whereas their chronological ages differed by a mean of over nine months.

Treatments:

Each group spent 40 minutes a day in class, five days a week, for approximately five months. The same rectangular 12-by-17 foot room was used by both groups, in turn, the equipment for one group being stacked away on shelves when the room was in use by the other group. Each group was in the charge of a different male graduate student.

1. Teaching Machine Approach. This approach followed a cafeteria principle. Several different machines were to be available at all times for the children to use as they wished. These were as follows:

a. The Story Machine: The basic machine was the START machine invented and developed by Martin Bender (Bender, 1965, 1966). In operation a paper tape bearing a printed story passes in view under a plexiglass window while the story is narrated aloud from recording tape attached to the paper tape. At certain words, the machine stops, displaying three printed words, each under a plexiglass key. By pressing the key over the correct word, the word is repeated aloud by the machine and the story continues. Pressing an incorrect key inactivated the machine and darkened the display, requiring that a reset button be pressed before another choice could be made. A series of familiar children's stories were programmed for presentation in this machine, and were changed at intervals of 7 to 14 days, as children appeared to lose interest in a particular story.

b. The Multiple-Choice Machine. Same machine as in a, but instead of presenting a continuous story, presented frame-by-frame discrimination tasks. In early programs, a model letter or word would be presented visually and aurally, with three visual alternatives from which the child had to choose the

Table 2.13.1 - Descriptive data on subjects in pilot study of teaching reading to two- and three-year old children

			ITPA Scores (Language Age)									
Subject No.	Sex	C.A.	Total	Subtests[1] 1	2	3	4	5	6	7	8	9
			Teaching Machine Group									
1	F	2-7	3-5	5-0	BN[2]	2-10	4-4	3-10	2-9	3-1	4-2	BN
2	F	2-10	2-11	3-4	BN	2-6	2-7	3-2	3-2	BN	3-1	3-11
3	M	2-11	4-3	5-2	4-5	4-8	2-3	3-10	4-2	3-10	4-4	3-11
4	M	2-11	4-1	4-7	4-9	3-6	4-0	5-8	3-2	3-6	3-7	3-11
5	F	2-11	(4-8)[3]	5-11			5-5				3-1	3-11
			Games Group									
6	M	3-3	3-0	2-9	3-0	2-6	4-0	3-10	3-6	2-4	2-6	3-11
7	M	3-7	4-4	3-10	4-1	4-5	3-4	6-7	4-7	4-7	3-11	4-2
8	M	3-8	3-11	3-6	4-9	4-2	4-4	3-10	2-11	2-9	4-10	3-11
9	M	3-8	4-8	4-9	5-6	4-8	4-8	5-8	3-6	4-7	4-2	4-2
10	M	3-11	4-2	4-5	3-0	3-6	5-1	4-1	5-0	4-7	3-3	4-10
Mean		3-3	3-11	4-4	3-9	3-8	4-0	4-6	3-8	3-6	3-8	3-10

[1] Identification of ITPA subtests: 1 = Auditory Decoding, 2 = Visual Decoding, 3 = Auditory-Vocal Association, 4 = Visual-Motor Association, 5 = Vocal Encoding, 6 = Motor Encoding, 7 = Auditory-Vocal Automatic, 8 = Auditory-Vocal Sequential, 9 = Visual-Motor Sequential.

[2] BN = below norms for subtest (minimum scorable age level is approximately 2-6, varying slightly from test to test)

[3] Testing terminated after three unsuccessful attempts to induce subject to continue. Total language age estimated on basis of four subtests completed.

matching one. Later programs presented sight vocabulary by pictures and spoken words, to which the correct printed word had to be matched.

c. The Voice Machine. Similar to above machines, but without audio capacity. Instead, the machine was equipped with a voice trigger so that the child had to speak his response into a microphone in order to make the machine advance. The machine was programmed to teach sight vocabulary. The program module was a three-frame sequence, in which the first frame presented a picture and a word together, the second frame presented the word alone, and the third frame again presented picture and word together. Thus the first frame served as the "text," the second as the test frame, requiring the child to remember the word without the picture cue, and the third frame provided confirmation or correction. Three words would be presented in a given program. After each word had been presented twice according to the sequence described above, the first or "text" frame of the sequence was eliminated and only test and confirmation frames were used. The machine would advance whenever the child spoke into it, regardless of what he said, and therefore initial demonstrations and instructions had to be given to the children to establish the practice of "saying the right word."

d. The Quiz Machine. This machine was functionally analogous to a juke-box. It presented a panel of 40 selector buttons. Pressing any button would cause the machine to locate and play back a predetermined word from a drum covered with recording tape. The selector buttons could bear either printed words or pictures. The recorded word would continue to play over and over as long as the corresponding selector button was held down. As used in this treatment, half of the buttons carried printed words and the other half carried pictures representing the same set of words but not in the same order. Although the children were at liberty to play with the machine as they wished, it was suggested to them that they start with a picture of their choice and then hunt for the printed word that said the same thing.

e. Other equipment. An electric typewriter was available at all times. When a child typed at it, an attendant called out the letter sounds as they were typed. If a child showed a continued interest in typing, he was later invited to copy first letters and then words. (This will be recognized as a crude duplication of some of the features of Moore's talking typewriter system.) Also available was a set of four-inch high letters cut out of masonite and corresponding grooved plywood blocks into which a letter could be fitted. A low chalkboard was also available.

Coordination of the learning devices. Since children were free to select the device they would work with and shift to another at will, the programming of each device had to be coordinated in such a way that switching from one to another would not produce interference and would as far as possible provide continuity in learning. The same basic sequence of letters and words was used in all the devices, roughly following that of Bloomfield and Barnhart's Let's Read (1961). There was enough difference in the detail of the sequencing, however, that generally a child who shifted from one machine to another could expect to encounter different words but ones that embodied the same phonic principles as the ones he had encountered just previously.

2. Social Games Approach. The aim in this approach was to engage the children in a variety of group games that the children would find enjoyable simply as play, but that would involve printed symbols in such a way that the children would be motivated to learn them in the course of their play. The following are some representative games:

a. Roulette. A large (almost four-foot diameter), vertically mounted wheel of fortune was used in a variety of matching and identification games. Cards could be fastened around the wheel. In the simplest game, individual letters of the alphabet were mounted and each child was given a card with a letter matching one of the letters on the board. After each spin of the wheel the child whose letter matched the one indicated on the wheel was entitled to give the wheel its next spin. Later additional requirements were added--that the child detect the match himself, and finally that he be able to name the letter.

b. Train. Six-inch wide planks were used to form a rectangular railroad track along which the children could walk imitating a train. Signs (stop, go, slow, fast, and choo-choo) were held up tell the engineer what to do.

c. Store. Child One would tell Child Two what letter he wanted to get and hand him a card bearing that letter. Child Two would then go to Child Three, the storekeeper, and ask for that letter (in the form of a masonite cut-out). Child Two would take the cut-out to Child Four, who would find the corresponding block and fit the cut-out into it, then give it to Child Five, who would deliver it to Child Two and then proceed as Child One before him.

d. Little Boy Blue. As the teacher read the nursery rhyme, "Little Boy Blue," various children would play parts of sheep, cows, and Boy Blue. Once children were familiar with the rhyme, parts would be assigned by handing children cue cards bearing the words, toot, baa, moo, or sleep.

As the above examples suggest, almost any game or activity could be stretched to include some content relevant to reading. As much as possible, the letters and words introduced followed the same sequence as those introduced into the teaching machine approach, although games such as Train and Little Boy Blue had their special vocabularies. A typical 40-minute session would include six or more games or activities, the number varying with the stability of the children's interest on a given day. Some non-reading activities were usually included for variety or a bit of large-muscle activity, the staples being marching to recorded music and a restrained form of tag. Each day also included the teacher's reading nursery rhymes to the children.

Data Collected:

The Illinois Test of Psycholinguistic Abilities was administered once after three weeks of school and again after four months of school. At the end of each daily session in both classes the instructor dictated detailed notes on the day's activities and each child's performance. These furnished the main body of data for clinical assessment of various aspects of each program.

2.14 Results and discussion

Feasibility:

(1) Teaching Machine Approach: All of the children showed some interest in all of the machines and devices used in this treatment. However, their response to the treatment was in marked contrast to the steadfast daily absorption that had been reported by Moore in his approach or by the inventor of the START machine, Martin Bender, in his home trials of the story programs. From the beginning the children seemed to be more responsive to social influences than to the devices themselves, and this tendency increased over time. Which machine a child chose to play with at a given moment seemed to be determined by which machine some other child was playing with, so that the teacher had to be constantly intervening in small disputes, managing the taking of turns, and inducing a child to undertake some second-choice activity when he could not have an immediate turn at the machine of his choice. This latter effort consumed much of the teacher's time, with the result that the children came more and more to depend on him to select their activities for them. Moreover, the children seemed

to need quite a bit of sustaining encouragement from him to keep them going on any machine. Evidently the devices themselves were not sufficient to maintain a lively interest, or at any rate they could not compete with social events. Frequently a child who seemed to be absorbed in work at one machine would leave it abruptly to rush over to another machine where a child was receiving attention from the teacher.

It would appear that the fault did not lie in the devices so much as in the way they were handled. The writer has observed children reacting the same way with the ERE ("Talking Typewriter") machine when it was merely sitting in a nursery school classroom rather than in a private cubicle. On the other hand, he has seen children in a Montessorri school taking turns working at the START machine with the same concentration that they give to the traditional Montessori materials. Evidently machine approaches, in order to be fully successful with young children, must be depersonalized, either through having the children work in physical isolation or through conditioning the children in such a way as to minimize competing social stimuli and make the children resistant to them. The teacher in the experimental class described here was greatly concerned with social relationships and tended to engage in much more social interaction with the children than was required. In this, however, he was probably much like the typical nursery school teacher. The kind of teacher behavior that would be necessary to reduce social stimulation in the class to the point where it did not obliterate the stimulus value of the machines would be much rarer among people attracted to early childhood education.

Of the several machines used, the story machine had by far the greatest sustained appeal to the children. At the same time, it was the most "social" of the machines, telling the child a story and occasionally addressing remarks to the child himself. Bender (1966), has found that the insertion of more direct instructions to the child in the story program to look at such-and-such, find such-and-such, etc., led to even greater attention, as indicated by a significant reduction in the number of incorrect responses. The Multiple Choice Machine, which was actually the same kind of machine but programmed to operate in a more conventional frame-by-frame manner, was less popular but it seemed to elicit the most real task behavior. With sufficient teacher encouragement and over the course of several days a child would occasionally work at a word recognition program until he could make errorless runs through it.

The Voice Machine was only occasionally attractive to children, but their behavior in working with it was such as to encourage further experimentation with such a device. Unlike the push-button machines, which rejected incorrect responses, the voice-actuated machine would advance no matter what the child said and would respond even to a noise. Thus it presented the danger of reinforcing incorrect responses. In practice, however, the problem was not the children's making careless responses, but their reluctance to make any response unless they were sure it was right. In order that they might advance through the program, the children had to be instructed to say, "I don't know" into the microphone if they were unwilling to guess at a word. They quickly caught on to the significance of the confirmation frames and took them as replies to the "I don't know" statement.

The Quiz Machine was broken down most of the time and was too unreliable when it did function to permit any judgments as to its appeal or effects. It should be noted that all of the machines existed in somewhat primitive versions. and were given to mechanical difficulties. Difficulties that caused a machine to stop functioning altogether were less troublesome than ones that caused it to make errors--to respond to a correct answer as if it were incorrect or vice-versa--for these tended to destroy the children's confidence in the feedback they were getting from the machine or, in some cases, to keep them from grasping the nature of the task altogether.

In general, the two-year-olds in this pilot study seemed to enjoy a machine to the extent that they could interact with it socially. They preferred the story machine to the multiple-choice machine. It did not just talk, it talked to them. They enjoyed shouting into the Voice Machine, but it did not give them much response in return. Anecdotal reports from users of the ERE ("Talking Typewriter") are replete with instances of children reacting to the machine as a personality, addressing remarks to it, chiding it for errors, and so on. Similar reports are obtained from people who have used tape-recorders in nursery school teaching.

Perhaps, then, Moore's notion of an autotelic responsive environment--an environment that enables the child to learn by being dramatically responsive to his actions--is somewhat off the mark as a rationale either for his own teaching machines or for others. If, for the young child, the machine functions not as a miniature environment but rather as a special kind of sociable entity, then the main differences between one machine system and another and between machines and live teachers may not be in their logic but in the kind of social situation they make possible. Moore's conception of a responsive environment precludes any learning except by discovery on the part of the child; for the environment can only react, it cannot deliberately teach. But if the essence of the machine is its social behavior vis-a-vis the child, there is nothing to say that the machine's role must be passive. It would be interesting to experiment with machines that were forceful, demanding, or competitive as against playful, subservient, or highly supportive in their relation to the child.*

None of the machine activities used in this pilot study commanded as much interest or effort as printing at the blackboard, tracing or coloring letters, or any of similar kindergarten-type activities that were imported from the social games treatment group.

(2) Social Games Approach: No matter what the game, it took the three-year-olds quite a bit of time and much teacher effort before they would play it according to stable rules. Eventually they could play and enjoy even such a complicated procedural game as "Store" (described in section 2.13). But if the game had limited utility, if it could serve to teach only a few special words or provide practice in some limited kind of letter identification, then it was hardly worth all of the time it took to teach the game itself. Thus the "Train" game was abandoned early because, although the children all liked it, it was apparent that it would take far longer to get them past the point where they were devoting all their attention to getting around the track, precipitating derailments, etc., and up to the point where they were attending to the printed directions, than the learning of the five words would merit. On the other hand, Roulette proved well worth the time spent in teaching the basic procedures, because once the game was learned it could be used continually to teach new material.

---

* I do not mean to imply in this discussion that the young child believes or should believe the machine to be alive, to have a man inside it, or any such animistic notion. We did not encounter any such beliefs among the children and could certainly have hastened to correct them had they appeared. The point is simply that young children do not seem to maintain as much social distance between themselves and non-human things as do adults, and thus more readily establish human-like relationships with anything that shows some modicum of responsiveness.

As time went on an effort was made to evolve new activities that permitted a greater amount of time to be devoted to learning content in proportion to the amount spent learning procedures. "Seat work" activities, involving tracing and copying letters proved to be appealing to the children, to the extent that they would spend ten minutes or more at it daily. Materials developed for this purpose were tried with the two-year-olds in the teaching machine group, where they were also readily accepted and became a regular part of the day's activities. Cutting and pasting variations were added, to provide a regular supply of kindergarten-type fare which had the most sustained appeal of any activity for the children.

The game that finally became the core of the program, however, was one in which each child was given a box in which he accumulated cards bearing the letters and words he could identify correctly. Each day he was tested on some of the cards and lost those that he could not identify, and then was tested on new cards that he had not yet acquired. The children could not be said to enjoy the game itself, but they did enjoy getting the cards and the game brought forth strong competitive efforts. This game was also introduced into the two-year-old group, where it was not taken so seriously but where it nevertheless seemed to call forth more effort than any other activities.

Thus, by a process of trial and error in searching for activities the children would enjoy and that would direct their efforts in ways relevant to reading, we were led away from games as such and eventually to activities that were indistinguishable from the conventional seat work and drill of elementary school instruction.

Reading Attainments.

No child in either group learned to read in his five months of training, nor did any come close enough to make it appear worthwhile to administer a standard reading test. It must be admitted, however, that we were so naive about standards of reading attainment that we discounted achievements of the children that would actually have been measurable. At least four of the ten children should have scored above the 1.0 grade level on word recognition and word sound analysis, the best probably scoring at about the 1.5 grade level.

The best index of the children's learning was provided by the letter and word cards they accumulated as learned. Since they were continually being retested and losing words they had forgotten as well as gaining new ones, the number of cards they maintained gave a fairly legitimate estimate of the size of their reading vocabularies. The size, of course, fluctuated from day to day, but by the end of the term it was about as follows: Subjects 7, 9, and 10 (referring to Table 2.13.1), had 20 or more letters of the alphabet, and approximately 20, 30, and 50 words respectively. Subjects 3 and 4 had about half the letters of the alphabet and about 15 words. The rest of the subjects were very unstable in their retention and could generally not remember more than a couple of letters and a couple of words consistently.

Only one child, Subject 10, showed any stable ability to sound out words from their spellings. Subjects 7 and 9 could occasionally do so to a very limited extent. Subject 10 was the oldest and most maturely motivated child in the group, although he was surpassed in total language age by four other children. What is perhaps most noteworthy is that all five of the children who showed some stable learning (Subjects 3, 4, 7, 9, and 10) started out with language ages above four.

Changes on ITPA.

Table 2.14.1 reports changes in ITPA scores between pretest and posttest for the 8 subjects available for retesting. These scores are calculated as a difference between standard scores, estimated to the nearest month by linear interpolation in the norm tables of the ITPA. The scores showed a tendency to decline. In fact, the mean loss of .46 standard score points for total score is significant at the .10 level (t = 2.09, d.f. = 7). Most of the loss is concentrated in the encoding tests, on both of which the subjects showed not only a relative loss for their increase in age but an absolute loss in raw score points. Since both of these tests are sensitive to the motivational state of the subject at the time of testing, they could easily reflect situational differences or examiner differences (different independent testers administered the pre- and posttests). It is evident, however, that the reading programs produced no gains on variables tapped by the ITPA.

One comparison of interest is that between the two information processing channels that figure in the ITPA subtests, the auditory-vocal and the visual-motor. The subtests so grouped may be considered as roughly approximating the verbal-nonverbal split found in other mental ability tests. Ignoring encoding, on the verbal subtests (1, 3, 7 and 8), the children in the teaching machine group showed a mean loss of .47 standard score points, while the children in the social games approach showed a mean gain of .26. On the non-verbal tests (2, 4, and 9), on the other hand, the children in the machine group gained an average of .16 while the children in the games group lost an average of .41 points. As Table 2.14.2 shows, this interaction between treatment groups and channels is highly significant and accounts for virtually all the variation in cell

Table 2.14.1 - Changes in ITPA standard scores after 4 months

Subject No.	Total	Subtest 1	2	3	4	5	6	7	8	9
					Teaching Machine Group					
1	-.70	-.99	1.48	-.10	-.30	-.88	-1.00	-.49	-1.32	-.10
2	-.46	.12	.35	.04	-.21	-.27	-1.35	-.03	- .74	-.06
3	-.42	-.58	-.06	-2.04	1.45	-.32	-1.98	.43	.99	.24
4	-1.16	-.99	-.61	-.56	-.53	-.89	- .60	-.39	- .94	.24
					Games Group					
6	-1.32	-.03	-.56	.50	-2.48	-1.03	-1.00	-.24	-1.37	-.94
7	.29	.65	.54	.00	- .19	-1.72	-1.11	.91	1.80	.09
8	.46	-.01	1.07	.61	-.53	.00	- .16	.90	.45	-.48
10	-.36	-1.46	1.55	.58	-1.93	-1.06	.84	.23	.57	-1.05
Mean	-.46	-.41	.47	-.12	- .11	-.65	-.86	.16	-.07	-.26

Table 2.14.2 - Analysis of variance on changes in ITPA auditory-vocal and visual-motor channel scores (excluding encoding) for two-year-old and three-year-old reading groups

Source	df	Sum of Squares	Mean Square	F	p
Treatments	1	.0256	.0256	.06	n.s.
Error	6	2.4058	.4010		
Channel	1	.0016	.0016	.03	n.s.
Treatment Channel	1	1.6900	1.6900	28.07	$<.01$
Error	6	.3613	.0602		

means. (If encoding scores are included the effects remain the same, though a little less extreme. The interaction is still significant beyond the .05 level.) Although treatments are confounded with age, there is no reason to expect age to be responsible for differential changes due to channel. The results therefore would seem to indicate that the highly verbal nature of the games treatment and the non-verbal, visual-motor character of the machine treatment were reflected clearly in ITPA changes.*

2.15 Conclusions

In a pilot study of two approaches to teaching reading to young children five two-year-old children were put through a five-month cafeteria style program in which they spent 40 minutes a day playing at liberty with a variety of teaching machines and other devices. Five three-year-old children were put through a program of the same length, but one built around learning games.

No child in either group learned to read, although three children in the games group and two who were initially in the machine group acquired some stable recognition of 10 or more letters and 10 or more words. For all of them, however, the only stable learning seemed to come from one activity--a variety of flash-card drill. Moreover, all of the children who showed stable learning had initial language ages of over four years.

Standard scores on the Illinois Test of Psycholinguistic Abilities showed a tendency to decline from pretest to posttest, mainly in the encoding areas. When tests were separated into visual and auditory channel tests, however, a significant interaction was found, the machine-taught group tending to improve on visual tests and decline on auditory and the social games group tending to improve on auditory and decline on visual.

---

*See Section 3.23 for a similar comparison of a verbal and a non-verbal treatment with older children.

This little study was of considerable value to the project as a whole because it pointed out at an early date some serious misconceptions which could have rendered much of the later work useless. It brought into serious doubt the assumption that young children could learn only through play-like activities. The effort to maintain a free play-like atmosphere and at the same time keep steering the children's activities toward learning objectives proved to be a considerable strain on the teachers and a source of unrest to the children as well. The work-type activities--that is, those in which a definite task was assigned to be accomplished--were much easier to handle and were the only ones that gave direct evidence of promoting learning. Moreover, they were accepted by the children in a matter-of-fact manner that was not always as enthusiastic as their acceptance of an exciting new play activity but that was much more conducive to productive effort.

Montessori's observation made 50 years before that young children liked to work appeared much more plausible after this experience. The question now, however, was what kind of work, how chosen, and how motivated? The experience of this pilot study also undermined confidence in the assumption that children's interest and natural curiosity provided a reliable guide to the kinds of activities that were most beneficial to them. In Montessori schools, children of two and three devoted much of their time to practicing household tasks such as sweeping, dusting, and setting table--and this was justified on the grounds that they appeared eager to master these skills. But it seemed to us that young children would eagerly strive for mastery of any skill that was within their reach and that the significant people in their environment had managed to convince them was important. Thus, the most popular activity of any that was hit on in the experimental classes was one that required the children to trace with crayons down the middle of outline letters. This was a task that presented some difficulties for children this age, but one where improvement came readily and mastery was possible. From both teachers and parents (who were exposed to daily bundles of colored-in letters brought home by the children), the children received evidence that this was a highly significant kind of accomplishment. Whether it was or not is an entirely different matter. In one line of attack on reading it might be and in another it might not. The children, certainly, were in no position to tell. The burden for determining what was significant clearly had to rest with the adults in the situation. It did whether they chose to assume responsibility for the choices or not.

Finally, the experience supplied no encouragemenet whatsoever for the popular belief that young children are highly motivated by a desire to explore and understand environmental events. With the teaching machine treatment it had been supposed that the children would be curious as to why some responses were accepted by the machine and others not or why some words came out of the machine instead of others. Moore's conception of the child as an active theory builder, as well as Piaget's notions of assimilation and accommodation would lead one to expect this. However, there was virtually no evidence of intellectual curiosity directed toward reading by any child in either group--no evidence of puzzlement or cognitive conflict, no evidence that discoveries were made or even sought.

The two- and three-year-old children we worked with were alert and superficially curious. They were eager to handle, inspect, and identify new things;

but having done this they were then ready to encounter something else. The interest that some older children show in figuring out the rules governing phenomena was, if not completely alien to them, at least too tenuous to serve as a motivational basis for an educational program.

What the children did show, as noted above, was an interest in mastering what they had been led to believe was important. They were quite receptive to indicators of what was important--not only from teachers, but from other children, from parents, and even from machines, if the machine behaved in a personal enough manner that it carried some authority.

2.16 Summary

Five two-year old children and five three-year-old children of average to above average language aptitude participated in five months of reading activities. The two-year-olds engaged mainly in free play with a variety of teaching machines and other manipulative devices; the three-year-olds engaged in teacher-directed group games designed to teach letter recognition, word recognition, and eventually some phonics. In both groups little visible learning occurred until an activity was introduced that amounted to conventional flash-card drill. Half of the subjects attained some stable sight vocabulary in this way, ranging from 10 to 50 words, and one child acquired some ability to decode novel words. Scores on the Illinois Test of Psycholinguistic Abilities indicated no relative gains in language-related abilities; but there was a significant interaction between treatment condition and changes in verbal versus non-verbal scores, children in the group games condition showing more favorable results on verbal tests while children in the teaching machine condition showed more favorable results on non-verbal tests.

2.2 The Effect of Free-Time Use of a START Teaching Machine on Reading Ability in the Kindergarten

Carl Bereiter and Martin Bender

2.21 Problem

In the preceding study one of the procedures used in the teaching machine approach with two- and three-year-old children involved the presentation of stories on the START teaching machine (see section 2.13 under "Treatments"). The machine narrated a story as it was presented visually on a moving tape. At response points the machine would stop until the child had pressed a panel over the correct printed word. This was the most popular of the machine procedures used, but the design of the study did not allow any assessment of its independent effects. The present study was designed to determine what, if any, learning would occur if the machine were merely made available to children to use as if they wished, without any attempt to supplement its teaching or integrate it with other curriculum content. Machines were simply placed in kindergarten classrooms so that children could use them during their free time if they wished, as they might use other free-play materials. A post-test administered to experimental and control classes provided the criterion of learning.

2.22 Design and Procedures

Subjects

The unit of sampling in this study was the class rather than the individual child. Ten experimental kindergarten classes were selected in five elementary schools in Urbana, Illinois. With the help of the Director of Special Services* matching control classes were identified for 9 of the 10 experimental classes (the unmatched class comprising a very heterogeneous group for which there was no comparable control class). The total sample represented all the public primary schools in Urbana, with the exception of one predominantly lower-class Negro school. Matching was judgmental, based on socio-economic status and "level" within schools. "Level" refers to the practice of assigning children to kindergarten classes within schools on the basis of judged maturity--a compound of age, intelligence, and social maturity. Whenever there was doubt as to the selection of an appropriate matching control class it was resolved in such a way as to favor the control condition--i.e., by choosing a control class of higher socio-economic level or maturity level than the experimental group with which it was paired. Two of the experimental classes and their matching controls were from one school, which was closed because of wind damage shortly before the end of the experiment, so that test data could not be collected. Thus the final sample consisted of seven matched experimental and control classes plus an extra experimental class.

---

* The writers are indebted to Gary Blade, Director of Special Services in the Urbana, Illinois, Public Schools for the valuable assistance in this and other phases of the experiment.

Criterion Test

It was assumed that the experimental effect, if any, would be small in comparison with experimental error. To maximize the likelihood of detecting an effect, the criterion test was deliberately based on words that had been used as response items in the teaching machine programs. The test, therefore, must be considered a simple test of learning and not of transfer or generalization. Part I consisted of 20 pairs of words. The subject was to circle the word in each pair that was read aloud by the examiner. The following pairs were used, the correct word being underlined:

it	*and*	*in*	but	just	*sat*	swim	*not*
big	that	had	*will*	*that*	ran	well	*it*
fast	*end*	*up*	big	*just*	is	*ran*	not
little	sat	with	*is*	will	*had*	*with*	and
this	let	*him*	can	him	*went*	end	*but*

Note that in each pair the first letter is sufficient to distinguish the correct word. Part II consisted of two sets of six picture-word matching items. The subject was to draw a line connecting each word with its corresponding picture. The sets of words were as follows:

bed
can
jump
knock
swim
hen

cat
man
witch
hill
pig
back

Procedure

One machine was placed in each classroom, at a spot removed from the main activity areas. On the first day an experimenter demonstrated operation of the machine to the class, after which there was no further intervention on the part of the experimenters. Teachers were instructed not to urge children to use the machine or to make it a subject of discussion with them, and to follow their normal curriculum, including whatever reading readiness activities it might entail. A project worker made the rounds of each class after it was ended each day, to record the number of correct and incorrect responses made during the session and the length of time the machine was in use that day, and also to make repairs and change programs as needed. Response data were collected by automatic counters in the machine's circuitry and time was recorded by an electric clock which operated automatically when a child was sitting on a bench that was part of the machine. The experiment went on for 14 weeks, beginning in January. A new story was put on the machine every two weeks, or sooner if daily time records showed a sharp decline in use.

The criterion test was administered during class time, as a group test, by an experimenter and two proctors.

2.23 Results

Mean scores for each class on the criterion test are presented in Table 2.23.1. Since the unit of sampling was the class, these means are treated as single, unweighted scores. The score for the unmatched experimental group was not used in the analysis of results. It may be noted, however, that although this class represented one of the least privileged groups in the study, its scores exceeded those of all but one of the control classes.

On Part I of the test, all but one of the experimental classes exceeded its matched control class. The mean difference is 1.71 points. The standard deviation of the differences is 1.68. This mean difference falls just short of significance at the .05 level (t = 2.44, d.f. = 6; t of 2.45 is required for significance at the .05 level).

On Part II every experimental class exceeded its control. The mean difference is 1.27 and the standard deviation of differences is 1.49. The mean difference falls short of the 2.45 value required for significance at the .05 level (t = 2.24).

On total score (unweighted sum of scores for Parts I and II), every experimental class exceeded its control. The mean difference is 2.99; the standard deviation of differences is 3.10. This mean difference is significant at the .05 level (t = 2.56).

Table 2.23.1 - Experimental and control class means on criterion reading test

Matched	Machine Class				Control Class				Difference		
Pair	N	Part I	Part II	Total	N	Part I	Part II	Total	Part I	Part II	Total
1	24	12.2	2.6	14.8	16	11.6	2.5	14.1	.6	.1	.7
2	25	14.4	4.3	18.7	No control class						
3	22	13.8	4.0	17.8	16	12.3	3.9	16.2	1.5	.1	1.6
4	22	14.9	6.5	21.1	23	10.0	2.0	12.0	4.9	4.5	9.4
5	19	13.9	3.7	17.6	17	12.1	2.9	15.0	1.8	.8	2.6
6	22	15.1	4.6	19.7	21	12.1	3.3	15.4	3.0	1.3	4.3
7	23	12.9	5.0	17.9	23	13.3	4.1	17.4	.4	.9	.5
8	18	16.6	7.8	24.4	23	16.0	6.6	22.6	.6	1.2	1.8

Thus, the differences, though small, were quite consistent in favor of the experimental treatment, resulting in a statistically significant overall mean difference, in spite of an effective sample size of only seven pairs. Had the analysis been based on individual scores rather than on class scores, the difference would have been highly significant.

Table 2.23.2 - School means for IQ, average time per pupil spent at teaching machine, and mean experimental group reading score

School	Mean I.Q. Grades 1-3	Mean Minutes per Pupil	Mean Score on Posttest
A	101.6	38.6	16.75
B	104.1	362.1	19.45
C	111.2	75.9	18.65
D	115.8	82.1	21.15

Table 2.23.2 presents some data on the relation between amount of machine use and criterion performance. Data are pooled for the two experimental classes within each school (the unmatched experimental class is included in school A). The mean time per pupil varied greatly, depending on the way teachers scheduled free time. In school B the schedule was arranged so that some children were always free to work on machines if they wished, whereas in school A only one period of free time was scheduled during which children had to compete for or take brief turns at use of the single machine available. Aptitude test data were not available on the kindergarten children themselves, and so the best estimate of intelligence level available was the mean IQ score for children in grades one through three tested the previous fall. There is an evident relation between mean IQ and criterion test scores, but in school B it is offset by the extra amount of time devoted by pupils to machine use. The relation between individual test scores and time spent on the machine would, of course, be more informative. But there was no practical way of obtaining individual data on machine use. Thus it is not known whether machine use was distributed more-or-less evenly over pupils or whether a small number of pupils accounted for most of the use. Daily time records did indicate, however, that machine use within the various classes remained relatively constant from week to week throughout the period of the experiment, except for reductions due to machine breakdowns.

2.24 Discussion and Conclusion

The results of this experiment do not permit any evaluation of the educational significance of the learning that took place. The predictive validity of the criterion test is not known. What was demonstrated, however, was that measurable learning did take place under very minimal conditions of intervention. The effect was observed in schools that varied from average to well above average levels of intellectual aptitude. Although it would require an extensive program of research to determine the kind of programming that would produce optimal effects on later reading performance, the fact that definite learning in the area of reading can be produced with a small investment in equipment and with no investment of teacher time suggests that the use of free-access teaching machines in kindergarten classrooms should be seriously investigated as a way of supplementing kindergarten education.

2.25 Summary

The main comparison in this study was between seven kindergarten classes in which a teaching machine was placed for free-time use and seven matching control classes which had no machines. The machine was of a type used in the previous study, which presented stories both auditorily and visually and which required the child to make occasional choices between printed words in order to keep the story going. On a recognition test consisting of words used in these choice-frames, every experimental class did better than its matched control class. The difference, using classes as the sampling unit, was significant at the .05 level.

2.3 Children's Preferences for High Versus Low Frequency Words

Carl Bereiter and Charles Summers*

2.31 Problem and Objectives

It seems to be taken for granted that beginning reading materials should be made as interesting for children as possible. This is a consideration in the choice of situations, plots, illustrations, and general format. Ordinarily it is not a consideration in the choice of words. Words are chosen according to principles of frequency of use, supposed ease of learning, or because they fit into some pedagogical scheme. And yet it is the words which the child is expected to learn. If all a child does is learn an interesting story by heart or retain the image of an amusing picture, the objectives of reading instruction have not been achieved and have indeed been subverted.

Certain reading materials for the popular market are claimed to be built around words which have some intrinsic appeal, usually auditory, for children; but to our knowledge these claims of interestingness have never been tested empirically. It is possible that the words were chosen because of their appeal to adults and that the same kinds of words are not appealing to young children.

The interestingness of a word could be based on any of a number of dimensions--on sound, appearance, the appeal of the thing it refers to, or upon its emotional connotation. It would be difficult to isolate such factors, because young children are usually not able to evaluate things in any but a global way. It is possible, however, to use their global judgments ("I like this one") to test hypotheses about factors that are involved in the appeal of isolated words.

---

* Deceased.

The present study was designed to investigate two questions:

1. Do young children have consistent word preferences of any sort? If children as a whole do not consistently prefer certain words over others, then the whole question of appeal of isolated words is likely to be a fruitless one to investigate.

2. Do children prefer words that occur relatively frequently in their own speech or do they prefer words that occur relatively less frequently? This question confronts the most popular strategy for choosing words for primary reading materials--frequency of use. The whole question of children's preferences for familiarity versus novelty or certainty versus uncertainty is under active investigation in experimental child psychology, but so far as we know it is not being investigated with respect to words.

Whether children will prefer novel or familiar stimuli seems to depend to a considerable extent upon the kinds of stimuli employed. In as yet unpublished research, J. McV. Hunt finds evidence that infants, when presented objects to look at over their cribs, shift from a preference for familiar objects to a preference for unfamiliar ones at around the age of three months. A study in progress by Sarah Johnson at the University of Illinois seems to be showing that a shift from preference for regular light patterns to a preference for irregular light patterns takes place between the ages of four and eight. Friedlander and Dixon (mimeograph) found evidence for a similar shift in comparisons between four and five-year-old nursery school children.

Clearly, it is impossible to generalize these results to children's preferences for words, but they suggest that familiarity or predictability may be an important variable in these preferences. The possibility seemed tenable, therefore, that in selecting words from among those used most frequently by young children, authors of beginning reading materials might be choosing words which are so lacking in novelty to young children that they might find them uninteresting to work with. This, in turn, could present motivational problems of some consequence; for even if story materials are interesting, the fact remains that most reading programs involve a considerable amount of work with the words in isolation, where the motivating effects of story context are lost.

2.32 Procedure

Word Sample: Populations of high- and low-frequency nouns and adjectives were defined on the basis of Rinsland's (1945) count of 354,000 running words recorded from the speech of kindergarten children. High-frequency nouns and adjectives were defined as those occurring 80 or more times; low-frequency nouns were those occurring 3 to 5 times; low-frequency adjectives were those occurring 3 to 8 times. Words of extremely low frequency were not used because of the likelihood that they would not be understood at all by most children and thus could not be used in meaningful textual material. These 4 populations comprised 185, 92, 36, and 39 words respectively. Thus the adjective populations were considerably smaller than the noun populations, reflecting their more limited occurrence in the speech of young children. Twenty words were randomly selected from each of the 4 populations. These are the words listed in Table 2.33.1.

Subjects: *S*s were 10 boys and 10 girls unsystematically chosen from a single kindergarten class in a small-town Illinois elementary school serving a village and surrounding farm region.

Experimental Design: A paired-comparison design was used in which each high-frequency word was paired once with each low-frequency word of the same grammatical class and a choice between the two made by one subject. This was accomplished by use of a 20-by-20 Latin square which assigned a different combination of the 20 high and low frequency words in each class to each *S*, so that all the possible combinations were exhausted.

Procedure: Words were printed in half-inch high letters on 1 1/2-by-3 inch cards, one word per card. Words were pre-arranged for each *S* in accordance with the design described above. *S*s were tested individually in a room apart from the classroom, and the testing was completed for all *S*s in one session to minimize inter-subject contamination. Words which *S*s were to choose between were arranged in two stacks, with right-left position of words balanced and randomized. The following instructions were given:

> Do you see these cards with words on them? I'm going to give you some of these cards to keep. They'll be your own word cards. You can take them home and do anything you want with them. I'll hold up two words and you can choose the word you want to have. (Hold up first two words.) Here are _____ and _____. Which word would you like to have? (If *S* points at the word.) Tell me which word you want, ______ or _____? All right this one is _______ (chosen word). You keep it. Now here are _____ and _____. Which word do you want to have?

Each S made 40 choices--between 20 pairs of nouns and 20 pairs of adjectives. On a random and balanced basis, half the Ss were given nouns first and half were given adjectives first. Motivation to execute the task and obtain the cards appeared very high, so high that children not included in the experiment had to be pacified when it was over by being given word cards for themselves.

2.33 Results

The design used in this experiment is an instance of double-sampling. Both words and subjects were sampled, and so the results can be analyzed statistically in terms of generalization to the population of words (what results would be predicted for repeated samples of words from the same populations?) and generalization to the population of Ss (what results would be predicted for repeated samples of children from the population of kindergarten children of which the present group is considered to be a sample?).

Table 2.33.1 presents data relevant to generalization of results to the populations of words. Since each word was paired once with each word of the contrasting frequency but the same grammatical class, the proportion of times that word was chosen may be taken as an estimate of the proportion of times it would be preferred to all words of the contrasting frequency type. All but three high-frequency nouns were chosen more than 50 per cent of the time in comparisons with the 20 low-frequency nouns. Overall, high-frequency nouns were chosen 66 per cent of the time. The standard error of this mean percentage is

Table 2.33.1 - Percentages of boys and girls (N = 10 each) choosing each word in paired comparisons with words in opposite list

High-Frequency Words	Boys	Girls	Total	Low-Frequency Words	Boys	Girls	Total
				Nouns			
Apple	70	60	65	Celery	20	50	35
Band	80	60	70	Chin	20	30	25
Barn	40	70	55	Coaster	20	30	25
Bat	40	10	25	Dominoes	20	50	35
Bed	70	60	65	Dynamite	40	40	40
Clay	50	90	70	Flavoring	30	40	35
Face	30	60	45	Galoshes	30	20	25
Farm	70	60	65	Gopher	40	30	35
Father	80	60	70	Hermit	40	20	30
Hen	40	50	45	Locust	30	20	25
House	60	70	65	Marshmallow	50	70	60
Monkey	100	70	85	Microphone	50	30	40
Morning	70	60	65	Nightgown	20	60	40
Pet	90	70	80	Parachute	80	40	60
Pony	80	100	90	Seashore	40	20	30
Sand	60	60	60	Stadium	60	10	35
Shoe	70	90	80	Stripes	20	40	30
Summer	50	70	60	Trailer	20	20	20
Wagon	90	90	90	Trolley	20	20	20
Window	60	80	70	Vessel	50	20	35
Mean Percentage	65	67	66		35	33	34
				Adjectives			
Beat	60	70	65	Attractive	40	50	45
Better	50	70	60	Brave	70	20	45
Black	70	90	80	Chilly	40	50	45
Blue	50	100	75	Dumb	40	20	30
Brown	50	40	45	Empty	30	40	35
Cold	50	30	40	Fierce	30	10	20
Fast	70	50	60	Foolish	10	20	15
Fine	70	60	65	Gay	20	40	30
Funny	80	80	80	Grey	40	60	50
Glad	50	50	50	Important	80	30	55
Good	90	60	75	Rough	20	40	30
Green	50	60	55	Safe	60	50	55
Hard	50	70	60	Slick	20	60	40
Last	50	60	55	Slim	30	50	40
More	60	100	80	Striped	50	40	45
Much	60	60	60	Thin	20	20	20
Nice	60	60	60	Tough	40	30	35
Sick	60	40	50	Vacant	50	20	35
White	40	60	50	Wide	40	30	35
Yellow	70	80	75	Wonderful	80	30	55
Mean Percentage	59.5	64.5	62		40.5	35.5	38

only 2.4 per cent, so that this percentage is significantly different from chance (50 per cent) at well beyond the .01 level of confidence.

High-frequency adjectives were chosen just as consistently (only two were not chosen more than half the time), but the average percentage is not quite so high--62 per cent. The standard error of this percentage is 2.6 per cent, so that this result also is significant beyond the .01 level.

When the data are looked at the other way, to determine how consistently children prefer high-frequency words, the mean results are the same. The average child chose a high-frequency noun 66 per cent of the time and a high-frequency adjective 62 per cent of the time. The errors of estimate are different, however, since in this case they are based on differences between children in percentage of high-frequency choices rather than on differences between words. The standard errors were found to be 3.07 and 3.56 for nouns and adjectives, respectively, giving results which are again significant beyond the .01 level in favor of high-frequency words. Consistency over Ss was not quite so high as consistency over words. Four children out of 20 did not show a preference for high-frequency nouns; five did not show a preference for high-frequency adjectives. No child chose fewer than 40 per cent high-frequency words on either list, however.

There was a tendency for Ss to choose the second word named. Fifty-seven per cent of noun choices were words presented second, and 55 per cent of adjective choices were of this type. The former percentage is significantly above chance at the .05 level ($t = 2.31$) but the latter is not ($t = 1.32$). Since this factor was balanced with the frequency factor, it cannot confound the results presented above, but can only serve to depress the effect of frequency. Some idea of what the effect of word frequency might be when unaffected by positional preference may be gained by considering only those words chosen in the non-preferred position. For these words it was found that Ss chose the high-frequency noun 72 per cent of the time and the high-frequency adjective 66 per cent of the time.

Although girls showed a stronger preference for both high-frequency nouns and adjectives than did boys, this difference fell far short of significance and could well be a chance effect.

2.34 Discussion and conclusions

Even with the small sample of children used, this study showed highly significant preferences on the part of kindergarten children for familiar or high-frequency words as opposed to less familiar or low-frequency words. Thus, if the interestingness of isolated words is of any positive value in beginning reading materials, these results indicate that the traditional practice of selecting vocabulary on the basis of frequency of use is on the right track.

Although findings which vindicate established practices are not very exciting, they are important in this case because established practices of word selection had not taken interestingness into account and there was some reason to suspect that familiar words would not be as interesting for young children as less familiar ones. Evidently kindergarten children have not reached an age at which words of high frequency are so familiar as to be lacking in appeal.

This raises an important issue for further research, however. Throughout the primary grades words for reading materials typically continue to be chosen from among those that are highest in frequency of use. The results obtained in this study cannot safely be generalized beyond children who are just beginning reading instruction. It is possible that in children of seven, eight, or nine a shift in preference toward less familiar words may occur. This could be determined by replications of the present study with children at successively higher age levels.

Methodologically, the present study holds promise for studies to test further hypotheses about the appeal of isolated words for children. For instance, the same design might be used with dual lists of words which differed in their use of certain sound patterns, to determine whether there are kinds of words which are appealing to children because of how they sound. With respect to concrete nouns, the question might be investigated of whether preferences for words are any different from preferences for the objects named. This could be done by a study which used sets of printed nouns for one experimental condition, as in the present study, and pictures representing the same nouns for another.

Sex differences in word preferences would also seem worthy of investigation in light of the often-heard claim that beginning reading materials have more appeal to girls than to boys. Although sex differences in preferences for particular words cannot validly be analyzed in this study because a given word was not paired with the same other words for boys and for girls, some of the larger differences that appear in Table 2.33.1 are suggestive of attitudes related to sex role; e.g., the words *parachute*, *stadium*, and *brave*, which were markedly more preferred by boys than by girls and the words *nightgown*, *face*, and *blue*, which were more preferred by girls.

2.35 Summary

Children's preferences for high-versus low-frequency words were tested by a paired-comparison method using 20 pairs of nouns and 20 pairs of adjectives having high versus low frequencies of occurrence in children's speech. Twenty kindergarten children were given randomly balanced pairings of the full set of words, with order balanced within pairs. For each pair *S* was asked to choose the word he wanted to have. For both nouns and adjectives, *Ss* chose more high-than low-frequency words, both differences being significant beyond the .01 level.

2.4 Four Approaches to Construction Activities in the Nursery School and Their Relation to Creative Problem Solving

2.41 Problem

It is tempting to avoid the issue of creative thinking in young children altogether. One can make a fairly good case for the claim that it doesn't exist, and thereby have done with the matter. To take up the issue, on the other hand, is to deal with the booming buzzing confusion that usually occupies any vacuum resulting from the absence of firm criteria. Creativity in young children is variously equated with disinhibition (the child who paints in smooth, sweeping strokes is creative; the one whose technique is arthritic is not), emotional self-expression, or childish innocence itself.

Creativity, in its common usage (which is the only usage intended here--we shall not try to endow the term with theoretical status), has two aspects. One is the aspect of originality, meaning not necessarily that the person does something that has never been done before, but that what he does is accomplished with a minimum of direction or example. Thus, at a trivial level it may be said that the child who places a rock on a newspaper when told, "Go do something so that the newspaper won't blow away," is behaving more creatively than the child who does the same thing having observed someone else do it or having been told just what to do. The other aspect is an esthetic one--what is done should strike us as particularly apt, clever, or beautiful.

Usually the second aspect is treated as the more definitive, since there are all sorts of trivial ways of measuring up on the first criterion. But it is the esthetic aspect that causes trouble in evaluating creativity in young children. What is one to make, for instance, of a child's painting that strikes adults as a delightful abstraction when, on questioning, the child reveals that he was trying to make the realistic representations? If one is going to call him an artist in spite of himself, then one might as well call him a comedian if his mispronunciations make people laugh. The characteristics of children's productions that strike adults as pleasing or ingenious, in other words, may not indicate any talents or behaviors in the child that are worth preserving or fostering. Note that this implies no criticisms of the adult esthetic judgments themselves. There is nothing wrong with judging children's paintings beautiful or their remarks profound, no matter how banal the children's intentions--but it is quite another matter to transfer these judgments of products to the child himself, the quality of his education and his future promise.

As a corollary to the above point, it seems futile to compare the creativity of children of different ages on the basis of esthetic judgments. The children's own standards and intentions can be expected to change with age, and the adult appeal of their productions can be expected to fluctuate also in a way that has no particular relevance to the children's progress. Thus, investigators who have attempted to apply uniform standards across age groups find a "fourth grade slump" in creativity (Torrance, 1962). This "slump" coincides with what Lowenfeld (1949) has noted as the time when children typically acquire an adult type of objectivity about their productions and realize that their drawings, for instance, don't resemble the intended objects nearly as much as they had supposed. It is easy to see how the dawning of self-criticism might produce a crisis in the child's creative pursuits, but the fact that what he produces at this time is less appealing to adults than what he produced when he was younger seems quite beside the point. It has no more operational significance than the closely related observation that, according to adult standards, children often go through a slump in physical beauty during the early teenage years and that many of them never recover.

The only sensible course, in light of the above considerations, is to ignore esthetic criteria altogether in studying creativity in children. And by esthetic criteria I mean not only criteria of artistic merit, but also criteria of cleverness and ingenuity, novelty, imagination, inspiration--in short, all the criteria according to which one might regard what a child had produced and say, "My, how creative!"

This leaves only the first aspect of creativity as a source of criteria--originality, or the child's success in accomplishing something without specific guidance. Although such criteria do not sufficiently delimit the domain of what popular usage calls "creative," they do at least include it. Moreover, they provide the opportunity for objective criteria which can, in the extreme, eliminate the problem of adult biases in judging children's performance. Thus, if a child does put a rock on a newspaper in response to the direction, "Do something so that the newspaper won't blow away," it may be difficult or fatuous to judge whether such an obvious response indicates imagination when performed by a five-year-old child, but it is at least possible to ascertain objectively whether or not the newspaper blew away.

Performance of this kind is usually labeled problem-solving rather than creativity, but the difference appears to be only in who sets the problem. The artist solves problems, but they are problems he sets himself as he proceeds with his work and the problems often cannot be formulated in such a way that they can be communicated unambiguously to others. The judgment of art therefore requires a well-developed esthetic tradition that ensures some common perception of what the problems are. Since no such common understanding can be developed with children and their productions, there is not much choice but to stay with objectively defined problems.

Construction or craft activities are among the few kinds where the young child may encounter complex problems that he has a reasonable chance of solving on his own. The child's effort, say, to build a bridge over a toy railroad forces him to deal with a number of engineering and design problems that are beyond him conceptually, but that he may nevertheless be able to work out practically. Such activities therefore provide a promising training ground for creative thinking and one which, unlike activities in the finer arts, permits some objective estimate of what the problems are and how successful a child is in dealing with them.

By far the most favored way of handling construction activities in kindergartens and nursery schools is that of allowing children to go about them freely--to set their own problems, in other words, and solve them with minimal help from the teacher. Such an approach, however, places considerable faith in the disposition of children to cope with rather than retreat from problems. It would not appear to be a very potent approach for the all too common child who merely "messes around" with materials in a purposeless and problem-free manner.

Assuming that learning takes place through coping with problems, construction activities could be improved by taking steps to see that children actually encounter problems. The simplest and most direct way would be to assign tasks that could be depended upon to face the child with problems. Such an approach runs the risk, however, of overwhelming the child with problems he is not equipped to handle. Alternatively, one can set construction tasks and then guide the child in executing them. This is the widely condemned but still very common method of "All right, now, everybody take hold of the blue strip of paper by the end, like this." It guarantees that children will go through the motions but seemingly eliminates problem-solving for everyone but the teacher.

The problem seems to be this: cast adrift, young children are likely to ignore or fail to handle problems; they need some help from a model, but when the teacher shows them what to do she is providing them with a model of the mechanics of construction, not of problem-solving. What might be better, therefore, is for the teacher to serve as a model of problem-solving, thinking aloud and demonstrating, so that the children can participate vicariously in the figuring-out process itself and learn something of it.

2.42 Objectives

This study was designed to provide a comparison of the effects of the four approaches suggested above on children's ability to solve construction problems independently, the four approaches being designated as follows:

1. Independent Construction, (FC), in which children are free to make whatever they wish as they wish.

2. Guided Construction, in which the teacher leads the children step-by-step through the construction of a specified object.

3. Independent Problem-solving, in which the children are assigned some object to construct but must work out the method themselves.

4. Guided Problem-solving, in which an object is assigned to construct, but the teacher serves as a model and guide in problem-solving, rather than leaving the children entirely to their own devices.

In accordance with the argument of the preceding section, the criterion was chosen to be as free as possible of subjective judgments of esthetic quality, imaginativeness, and the like. Children were to be tested by constructing specified objects (not previously constructed) as in condition 3, above. The criterion chosen was the simple one of the extent to which the child's production constituted fulfillment of the assigned task, in function and appearance. Thus, if the task were to construct a bridge, the criteria would not include cleverness or beauty, but simply such points as does it look like a bridge? does it span a distance? could it function (in miniature) as a bridge?

The four approaches, as labeled, differ on two dimensions. It was predicted that the guided approaches would result in better criterion performance than the independent approaches and that the approaches involving problem-solving would result in better performance than those lacking definite focus on problems. Thus the fourth approach, Guided Problem-solving, was expected to show the best results of the four, being a combination of the two favored characteristics.

2.43 Method and Procedure

Subjects: Subjects were children in four nursery school classes, all largely populated by children of university faculty members and graduate students. Classes A and B (n = 17 and n = 20) were the four-year-old classes in a cooperatively owned but staff-operated nursery school, Class B consisting of the older fours or those judged by the teachers to be more mature. Class C consisted of the four-year-old group (n = 21) in a cooperative nursery school. Class D (n = 20) was from the same school as Classes A and B, but consisted of the older and more mature of two classes of three-year-olds.

Treatment: Each class was divided into four treatment groups that were conducted simultaneously. Training for all groups consisted of six daily half-hour sessions, following which there were three daily testing sessions, also of half-hour length. Each child in every group was provided with an identical large assortment of construction materials that included such things as styrofoam balls, pipe cleaners, colored sticks, paper cups, sponges, dictaphone belts, paper straws, wire, modeling clay, scissors, and paper. After each training or test session materials were replaced to maintain a constant inventory. Training procedures were as follows:

Group IC (Independent Construction) had no assigned tasks, being allowed to construct anything they wished each day. Experimenters were permitted to give technical assistance when it was requested, but were otherwise restricted to offering encouragement, suggesting that the child try new projects or new materials, and supplying liberal praise for accomplishments.

The other three treatment groups followed a fixed sequence of construction tasks, one task each day in the following order: 1) scarecrow, 2) wagon, 3) kite, 4) cage for lion, 5) bridge, 6) giraffe. Group GC (Guided Construction) was guided through the construction of each object, following predetermined plans that were intended to embody novel uses of materials and interesting structural ideas.

Group IP (Independent Problem-solving) was simply assigned the task with no instructions or advice as to how to do it. Experimenters were limited to the same kinds of activity as in Group IC, except that they were directed to encourage a child to keep trying if he showed an inclination to give up on a task.

To Group GP (Guided Problem-solving) the task was presented as one that experimenter and children would have to work out jointly. Although the experimenter had the predetermined plans used with Group GC to fall back on, he was instructed to make as much use as possible of the children's ideas. The experimenter's function was to call the children's attention to problems and possibilities of elaboration and improvement as a construction project went along. Only if the children failed to come up with adequate suggestions of their own was the teacher to offer definite suggestions.

Testing: After six daily sessions administered according to the treatment descriptions given above, all groups were given three additional daily sessions under the IP (Independent Problem-solving) condition, in which they were told what to construct but given no guidance in how to do it. The tasks were a boat, a rattle, and a bird house. These represented the criterion tasks according to which the effects of the treatments were judged. These criterion tasks were selected at random from an original set of nine, the remaining six constituting the training tasks.

Each object produced during test sessions was photographed in black-and-white along with a coded label, and jumbo-sized prints made. The photographed objects were then rated independently by two raters, one an advanced undergraduate in architecture, and the other in industrial design. The raters had had no previous contact with the experiment and pictures were coded in such a way that the raters could not identify treatment groups, nor even which objects were made by the same child, thus eliminating any halo effect. Rating was done on a six-point scale of degree to which the product represented fulfillment of the task assigned. Raters were instructed to judge on the basis of both function (for instance, could the bird house house a bird) and figural resemblance, but not esthetic appeal or novelty.

Teacher-experimenters: Four teacher-experimenters were responsible for administering the training and test tasks. All were undergraduates--one male art student, two female students in elementary education, and one female student in communications specializing in children's television. All had had extensive experience in working with young children, though not as nursery school teachers. Each teacher was instructed in each of the methods and was closely supervised during the experiment itself to see that the assigned method was being followed. A given teacher-experimenter would work with the same group of children through an entire series of six training and three testing sessions.

Experimental design: The design of the experiment was somewhat complex in order to control for teacher and class differences. The four classes, four teacher-experimenters, and four treatment conditions formed a Latin square. Thus, every treatment was administered once in each class and once by each teacher-experimenter. In addition, however, the four treatments formed a two-by-two factorial design, one factor being guided versus independent activity (GC and GP versus IC and IP) and the other being emphasis on construction versus emphasis on problem-solving (GC and IC versus GP and IP). There were also two within-person factors--the three test tasks and the two independent raters.

2.44 Results

A five-way analysis of variance on product ratings was carried out, the factors being identified as follows:

Between subjects

A Nursery school class (4 categories)

B_1 Guided versus independent activity (2 categories)

B_2 Construction versus problem-solving emphasis (2 categories)

Within subjects

C Criterion task (3 categories)

D Rater (2 categories)

A fixed-effects model was applied throughout, even though factors A, C, and D could all be treated as random effects, because of the technical difficulties in applying a mixed model to a design of this size. A choice had to be made between including nursery school class or experimenter as a variable in the analysis, since their interactions were confounded in a Latin-square design. Preliminary analyses showed both factors to have significant main effects, but it was decided to include nursery school class on the grounds that its interactions with other factors might be more interpretable. As a result of photographic failures and absences, 74 of the 444 possible scores were not obtained. However, these missing data were scattered in such a way that at least two scores were available for each subject. Missing data were accordingly filled in by the method of marginal mean estimates suggested by Winer (1962, pp. 281-283), a method which should be expected to gain in accuracy with the number of factors involved. Error degrees of freedom were adjusted to reflect data loss. All four-way and higher interactions were pooled with their respective error terms, except for the B_1B_2CD interaction, which was pertinent to the interpretation of treatment effects.

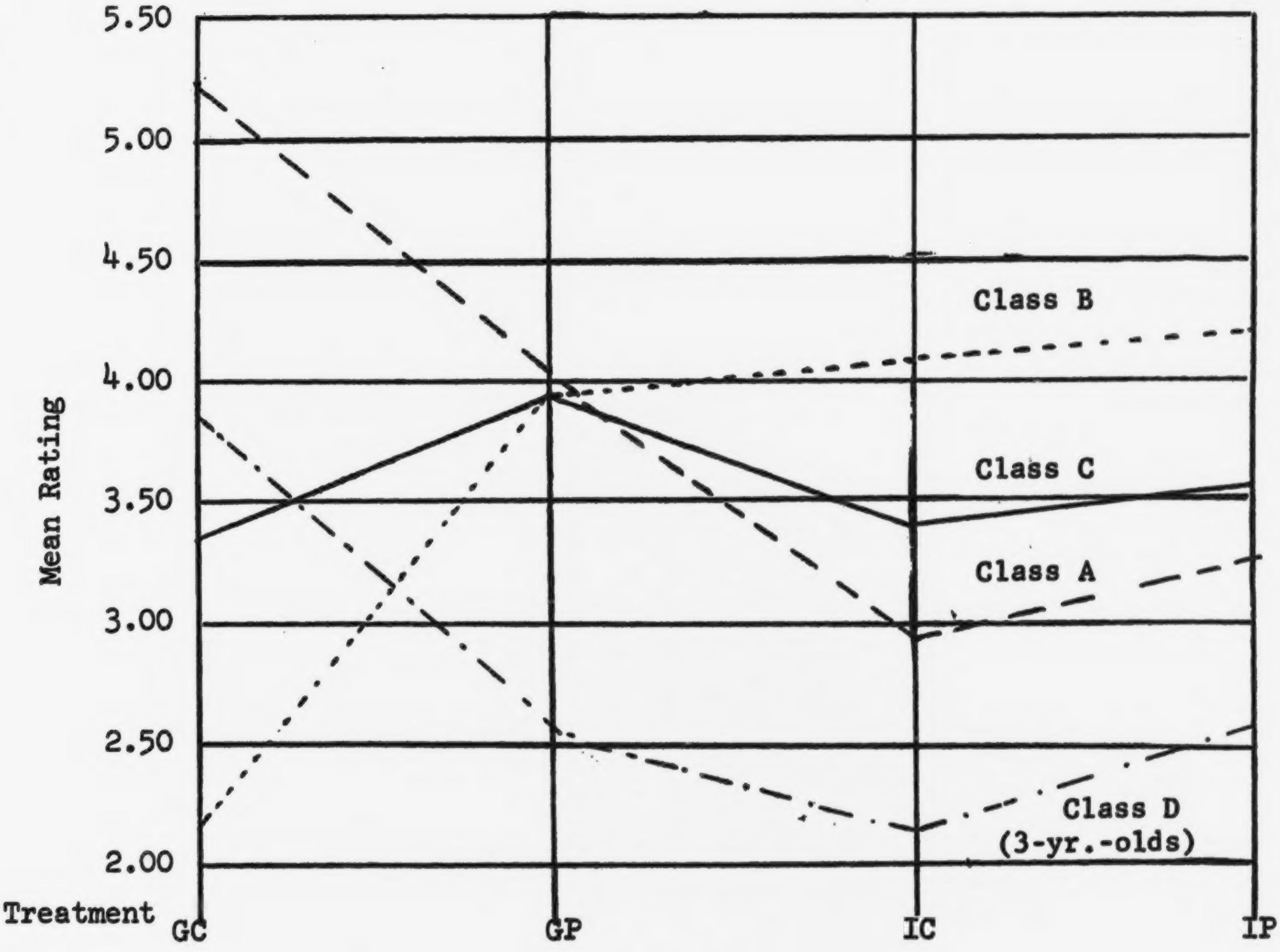
Table 2.44.1 presents the results of this analysis of variance. There is a highly significant A effect, which merely reflects the inferior performance of the three-year-old group compared with that of the other three groups composed of four-year-olds (a mean of 2.72 compared to means of 3.55, 3.88, and 3.75 for the other groups).

Table 2.44.1 - Five-way analysis of variance on nursery school class, two treatment dimensions, task, and rater

Source	df	SS	MS	F	P
Between Subjects	73	604.95			
A (Nursery School Class)	3	93.05	31.02	5.84	$<.01$
B_1 (Guided-vs.-Independent)	1	2.4.49	24.49	4.61	$<.05$
B_2 (Construction-vs.-Problem-solving)	1	.36	.36	<1	n.s.
AB_1	3	87.42	29.14	5.49	$<.01$
AB_2	3	23.03	7.68	1.48	n.s.
B_1B_2	1	7.22	7.22	1.36	n.s.
AB_1B_2	3	51.33	17.11	3.22	$<.05$
Subjects within groups	58	308.05	5.31		
Within Subjects	296	892.28			
C (Task)	2	26.36	13.18	2.65	$<.10$
AC	6	21.91	3.65	<1	n.s.
B_1C	2	18.15	9.06	1.83	$<.25$
B_2C	2	9.59	4.80	<1	n.s.
AB_1C	6	48.55	8.09	1.63	$<.25$
AB_2C	6	38.86	6.48	1.30	n.s.
B_1B_2C	2	7.42	3.71	<1	n.s.
Pooled error	85	422.33	4.97		
D (Rater)	1	8.38	8.38	4.25	$<.05$
AD	3	8.56	2.85	1.45	$<.25$
B_1D	1	.91	.91	<1	n.s.
B_2D	1	5.06	5.06	2.57	$<.25$
AB_1D	3	2.24	.72	<1	n.s.
AB_2D	3	4.38	1.46	<1	n.s.
B_1B_2D	1	2.80	2.80	1.42	n.s.
Pooled error	61	119.94	1.97		
CD	2	22.16	11.08	11.19	$<.01$
ACD	6	13.33	2.22	2.24	$<.05$
B_1CD	2	7.24	3.62	3.66	$<.05$
B_2CD	2	2.26	1.13	1.14	n.s.
B_1B_2CD	2	4.52	2.26	2.28	$<.25$
Pooled error	97	96.33	.99		
Total	369	1497.23			

The B_1 effect (guided versus independent activity) is significant at the .05 level, indicating superior criterion performance for the guided groups (mean of 3.73 versus 3.25). The absence of any significant B_2 effect or B_1B_2 interaction would suggest that the differences among treatment groups are wholly attributable to effects of teacher guidance versus the lack of it. The situation is complicated somewhat, however, by the presence of significant AB_1 and AB_1B_2 interactions. These interactions are graphed in Figure 2.44.1.

Figure 2.44.1 - Between-group Interactions



In Class C, the differences between treatment groups are small but indicate a B_2 effect--problem-solving (IP and GP) superior to construction

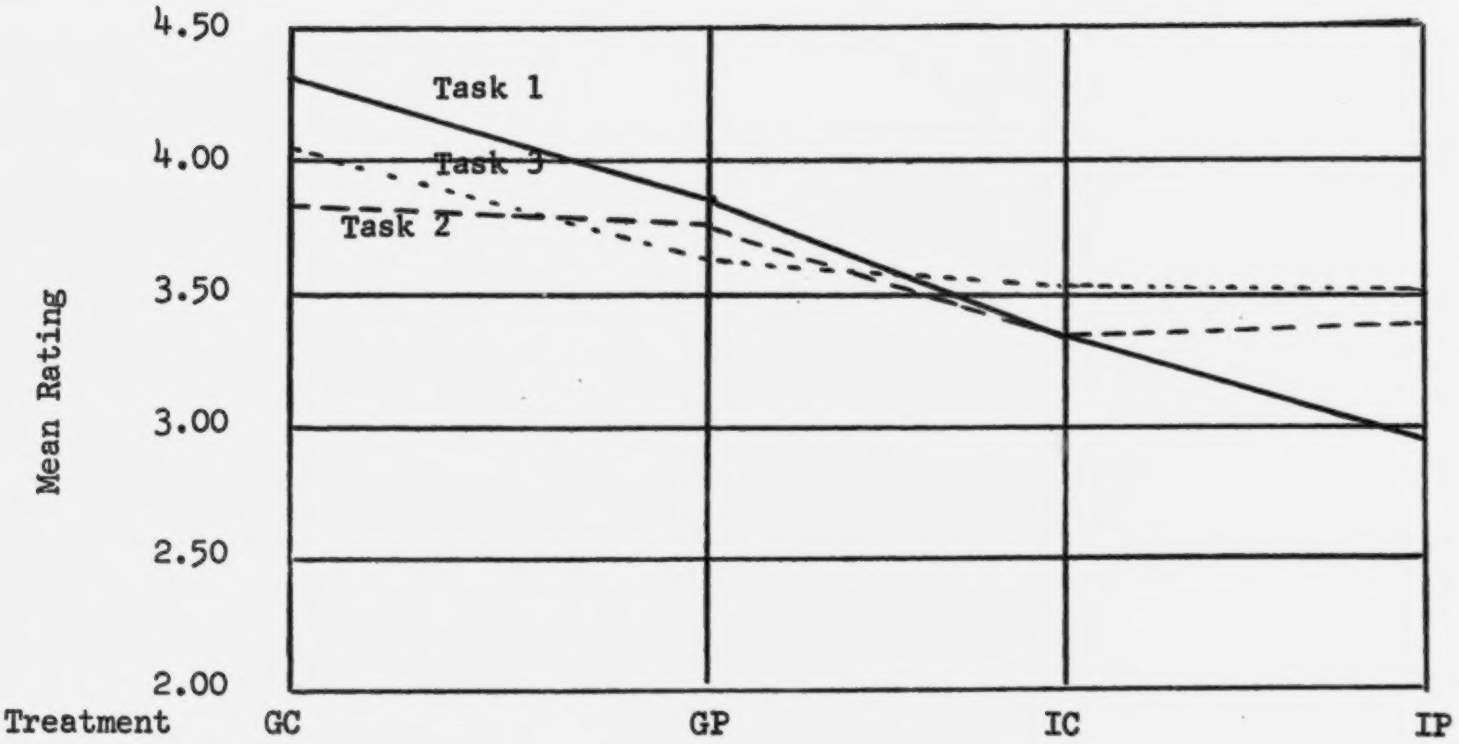
(IC and GC). Class A on the other hand shows a B_1 effect, the guided groups markedly exceeding the independent groups in criterion performance. Class D (the three-year-olds) reveals higher scores for Guided Construction than for any of the other treatments, while class B shows just the opposite. Thus the treatment effects are not at all clean.

Turning to within-person effects, a significant difference between raters (factor D) was obtained, and the difference between mean scores for the three test tasks (factor C) approached significance. These effects are of no interest however, and neither factor taken singly interacted significantly with the between-subject factors. A highly significant interaction between factors C and D was obtained, however, indicating that the two raters differed in the relative scores they assigned to the three kinds of criterion task objects. Moreover, this effect was involved in significant three-way interactions with factor A and factor B_1. Figure 2.44.2 illustrates the latter interaction. It is evident that Rater 1, shown in the left-hand graph, was quite consistent in the scores given to treatment groups on all three tasks, whereas Rater 2 showed little consistency--for instance, groups GP and IP each received the highest score on one task and the lowest score on another.

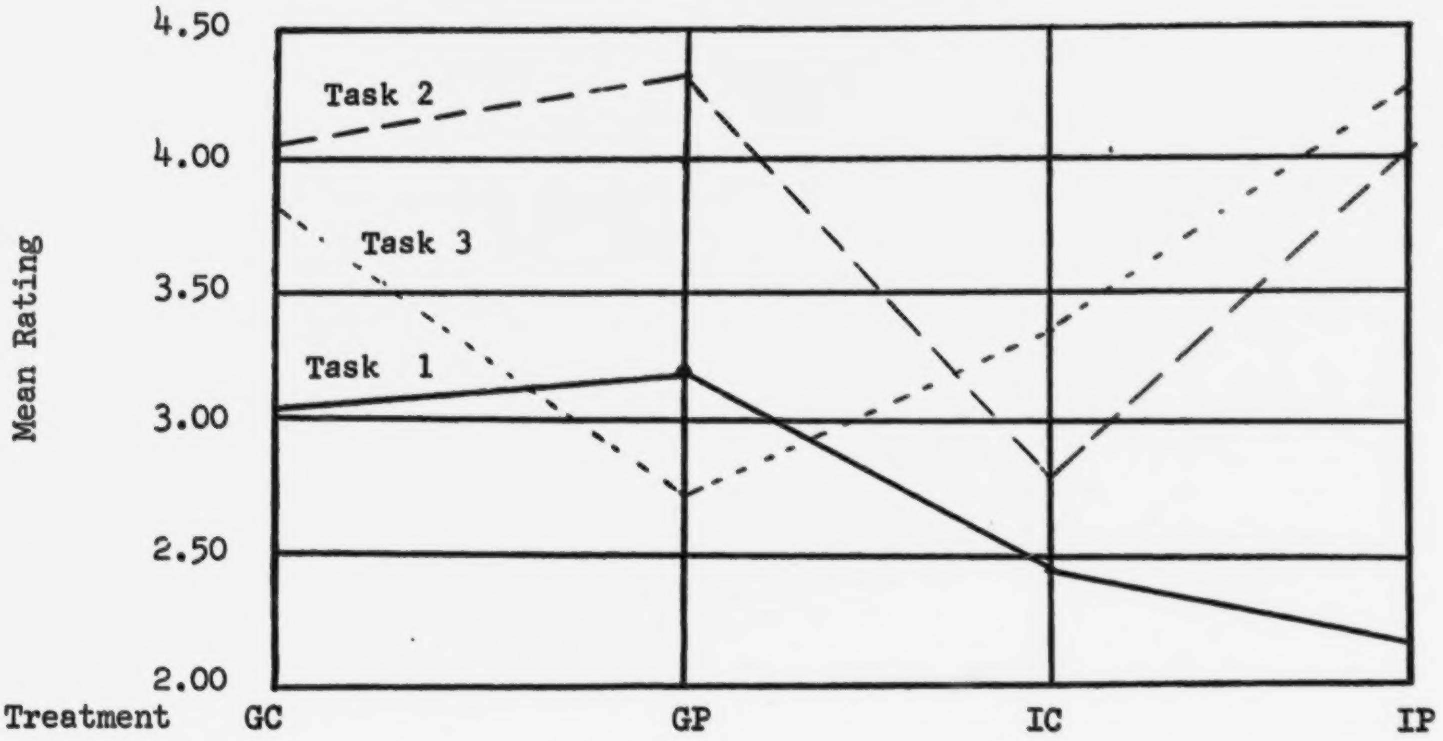
This interaction reflects on the reliability of the scorers. It will be recalled that the objects to be rated were coded in such a way that scorers could not tell whether different task objects came from the same treatment group or individual. Thus the consistency shown by Rater 1 could not have arisen from bias or failure to make independent judgments. Rather, it would indicate that his ratings were

Figure 2.44.2 - Task-by-rater-treatment Interaction

Rater 1



Rater 2



sufficiently valid that they reflected the same group differences from task to task. In the light of this consistency over measures, it would appear that the scores given by Rater 2 reflect erratic judgments or judgments on dimensions not common to the three tasks. The discrepancies between the two raters are further indicated by product-moment correlations between their ratings. For the separate tasks, the agreement is fairly good considering the ambiguity of the dimension being rated-- r = .47, .68, and .61 for boat, rattle, and bird house, respectively. But the sums of their ratings over the three tasks correlated only .49, whereas they would ordinarily be expected to correlate more highly than the separate task scores.

The above findings made it seem advisable to examine the results for Rater 1 separately. Only the between-person factors were examined, since it was already evident that there would be no significant effects related to task for Rater 1. This analysis is reported in Table 2.44.2. The results are essentially the same as for the five-way analysis of variance reported above, except that the B_1 and AB_1B_2 effects are substantially strengthened.

Table 2.44.2 - Reanalysis of between-person effects using scores from Rater 1

Source	df	SS	MS	F	P
A (Nursery School Class)	3	44.03	14.68	6.07	$<.01$
B_1 (Guided-vs.-Independent)	1	17.40	17.40	7.19	$<.01$
B_2 (Construction-vs.-Problem-solving)	1	1.36	1.36	<1	n.s.
AB_1	3	35.52	11.84	4.89	$<.01$
AB_2	3	15.62	5.21	2.15	$<.25$
B_1B_2	1	1.59	1.59	<1	n.s.
AB_1B_2	3	78.51	26.17	10.81	$<.01$
Subjects within groups	58	140.55	2.42		
Total	73	334.58			

Two final analyses of between-subject differences were carried out that attempted to deal with the problem of missing data by eliminating portions of the design rather than by estimation. In both analyses subjects were retained only if complete data were available on them. The first was a class-by-treatments analysis which, by using data from only the first two test tasks and by eliminating Class B altogether, retained 49 of the original 74 subjects. This analysis revealed no significant treatment effect ($F < 1$), but a Class effect and a Class-by-Treatment interaction both significant beyond the .05 level. The interaction, again, was similar in pattern to that shown in Figure 2.44.1. The other analysis used only data from the third test task and permitted use of 65 of the original 74 subjects, but eliminated one cell (the IC treatment group of Class C), thus breaking up the Class factor. A one-way analysis of variance revealed no treatment effect ($F = 1.64$, $df = 3.61$).

2.45 Discussion

There was an overall tendency for children who had received direct guidance in constructing objects to do better on the construction of test objects than those who had received only encouragement and praise. The effect of the direct "do it this way" type of exercise was especially pronounced with the youngest children. These findings cannot be dismissed, however, with the old slur of "teaching for the test." If the charge is applicable to any treatment, it is the Independent Problem-solving condition which was identical to the testing condition. If the children in the guided treatment conditions were to apply anything they had learned to the test tasks, it would have to be through transfer of principles or techniques to new problems, not through the simple repetition of learned patterns.

Most likely, what was transferred in the guided conditions was a superior repertoire of material uses to draw upon in construction tasks. Children in the unguided conditions tended to stick to a few materials during all their training sessions, even though the teachers praised any use of different materials. In the Guided Construction condition, on the other hand, constructions were planned so as to assure the children of experience with a variety of material uses. This was also true to some extent in the Guided Problem-solving condition, which in addition directed attention to different materials through the often-used question, "What could we use for such-and-such?"

Observations of the four replications of each treatment suggested that motivation was generally highest in the directed training groups, although this high level of motivation was quickly lost when teacher direction was withdrawn in the test sessions.

Independent Problem-solving was the most difficult condition to manage, and thus it presented difficulties as a testing condition as well. Children were often inclined to give up and say they couldn't do it as soon as a problem was posed. In the Independent Construction condition, children were generally enthusiastic for the first few sessions, but then became bored as they kept making the same things over and over. Suggestions that a child try something different were usually ineffectual, since the teacher was barred from suggesting anything specific.

Guided Construction was the most uniformly successful treatment from a procedural point of view, as one might expect it to be, considering its continued widespread use by teachers over the objections of authorities. Some of the children were unaccustomed to taking direction in their work, the nursery school classes being all of a somewhat liberal bent, but they came into line quickly when they saw that failure to go along with directions resulted in failure to bring forth the product that others in the group were so proud of. If nothing else, Guided Construction seems like a good method of conformity training.

Guided Problem-solving proved to be a difficult treatment for the teacher-experimenters to carry out. It is, after all, a "discovery method," and shares with others of this family the placing of a heavy burden on the teacher's resourcefulness, patience, and responsiveness to subtle cues from students. Only one of the teacher-experimenters struck the writer and his colleagues as being fully successful in bringing the method off. In her group an interesting transformation took place. She succeeded in getting the children intensely engrossed in each day's construction problem, mostly through example, and used questions to lead them on to productions that were striking to the children who made them as well as to observers. By the third session, however, the children began to take on much of the question asking themselves. Eventually the sessions came to resemble the Independent Problem-solving condition, the teacher-experimenter doing little more than introducing the problem and the children proceeding from there to a lively discussion of how to go about constructing the problem object. The difference, however, was that the children worked as a group, whereas in the Independent Problem-solving condition they worked independently except for copying. This made a difference when it came to the test tasks, where the children were not allowed to work out constructions cooperatively. This exemplary group, which was the GP group in Class C (see Figure 2.44.1), performed well but not at all outstandingly so on the criterion tasks.

The other Guided Problem-solving groups, though not so successful in realizing the intent of the method, were nevertheless saved from disaster by the provision that if the teacher-experimenter was not successful in drawing solutions from the children, he should allow them to participate vicariously in his own solving of the problem. This meant, in effect, a partial regression toward guided construction. The availability of this option is noteworthy in constructive or creative activities, since it is not usually available in other activities. In teaching mathematics, for instance, when a teacher's efforts to lead the children to discover a principle fail, her alternatives are either to turn guided into independent problem-solving with useless injunctions, such as, "Come on, think!" or else to retreat humiliatingly to direct instruction, saying in effect, "All right, since you were too dense to figure it out, here's the answer." For the teacher to think the problem through aloud herself and discover the solution, allowing the children to participate vicariously in the adventure, will be quickly spotted as the fraud it is. In creative activities on the other hand, the teacher can be authentically thinking aloud and at least trying to hit on a new solution, even though she has been over the same ground many times before.

The phenomenon of children in the Guided Problem-solving group either advancing to Independent Problem-solving or regressing to Guided Construction suggests, however, that the ideal approach includes all four of the treatments included in this study, but in a certain sequence rather than combined. Thus Guided Construction would be the logical starting point, especially with younger children, so that they could acquire techniques and familiarity with materials. Gradually the treatment could shift to Guided Problem-solving, in which the children became more involved in the planning, the teacher working out construction ideas with the class rather than beforehand. As the children began to acquire the ability not only to answer questions related to construction problems but also to ask them, the teacher could take a less active role, and the treatment would shift to Independent Problem-solving. Finally the children might be able set their own tasks, in which case they would have worked their way up to Independent Construction, but having acquired what it took to be successful at it along the way.

One final suggestion from this study is that creative problem-solving, however it is taught to young children, can probably best be taught to them in groups. Only in groups is the question-and-answer process that moves problem-solving along carried on publicly enough so that all may learn from it. The chance of success is also greater and the need for sustained effort and concentration is not quite so great as it is for the child working alone. The idea that creative thinking must be a solitary enterprise is probably true as regards complex productions--it is hard to imagine brain-storming as ever having much utility beyond the discovery of angles and gimmicks. But with young children complexity is never at a very high level anyway, and working in groups could be expected to have the effect of adding to rather than reducing the complexity of the aggregate product.

2.46 Summary

Groups of nursery school children were given six half-hour sessions of construction activities, using a large assortment of construction materials. Following this were three test sessions in which the children were assigned three construction tasks to be carried out independently without help or suggestions. The tasks were to construct a boat, a rattle, and a bird house. Photographs of the products were given blind ratings on the degree to which the product represented fulfillment of the task assigned (not on originality or esthetic value). A total of 74 children were involved in four replications of four treatments: Independent Problem-solving, in which daily tasks were assigned as in the test condition; Independent Construction, in which no tasks were assigned but the children were merely encouraged to make something different each day; Guided Problem-solving, in which teacher and children work out the solution of construction problems cooperatively; and Guided Construction, in which the teacher directs the children through a predetermined series of steps in the construction of prescribed objects. The analysis of ratings provided for the separation of the two treatment factors, guidance versus independence and problem-solving versus construction. Only the first factor showed any significant effect, and this was in favor of guidance over independence. There was no interaction between treatment factors; thus Guided Problem-solving, which was expected to have greater transfer value to the test situation, produced no better results than Guided Construction, which is popularly believed to discourage creativity.

2.5 Teaching Conservation of Substance

David W. Brison[1,2]

2.51 Prefatory note (by Carl Bereiter)

Although from a common-sense educational point of view learning to read is the intellectual attainment par excellence of the child during his second five years of life, to the many followers of Jean Piaget, the major intellectual advance in these years is signaled by the acquisition of a peculiar concept called conservation of substance. Conservation of substance refers to the fact that things do not increase or decrease in the amount of matter they contain when they undergo changes in shape or distribution.

The differences between children who have and who don't have this concept are often quite striking. Start with two balls of clay that the child acknowledges are equal in amount and mold one of them into a star. The child who lacks conservation may state very confidently that the star contains more clay and explain that this is because it has points on it which the ball does not. A year later the same child may state just as confidently that the ball and the star contain the same amount of clay and explain that they had the same amount before and that you didn't put any on or take any away.

Still, what is so important about this concept that should make it stand out above all the others that children commonly acquire between the ages of five and ten? To Piaget it is important because in order to grasp this concept (or to conserve, as it is put more operationally), the child must have accomplished most of the steps involved in progressing from a pre-logical stage of thought to the stage of concrete logical thinking. Conservation

---

[1] Now at Ontario Institute for Studies in Education.

[2] Dr. Luther Pfluger, Principal of the Hawthorne Elementary School in Oak Park, Illinois, was extremely helpful in providing the setting for the data collection.

involves reversibility (mentally reversing an operation to determine whether things return to their original state) and compensation (dealing simultaneously with two variables that change reciprocally to maintain some function at a constant value). It also requires some concept of atomism, of matter composed of small particles that can be rearranged without altering their number. The non-conserving child, according to this analysis, is dominated in his thinking by immediate perception and by the particular stimulus dimensions that he happens to focus upon. He cannot, therefore, successfully integrate past knowledge with present perceptions nor deal with relations involving more than one dimension at a time. Thus he is severely hampered in his efforts to understand the flux of states and relations in the world about him until he begins to "conserve" (Piaget, 1960).

If conservation is so important, then speeding up its acquisition should have far-reaching effects on the child's conceptual development. Unlike reading, however, it is not the acquisition alone that is important but also the subordinate learnings. With reading it is only the eventual ability to read that counts and the subordinate learnings, such as letter recognition and phonics, are of no particular value except as components of reading. If an efficient way of teaching reading were found that could circumvent these subordinate learnings, nothing would be lost. But conservation of substance in itself is not a concept of overwhelming importance for a child, and to teach it by some shortcut that circumvented the acquisition of reversibility or compensation would be to lose the point of Piaget's scheme of intellectual development.

In some pilot work on teaching conservation of substance to intellectually advanced kindergarten children, the writer had observed that merely teaching the children to predict correctly which way a balance would tip or whether it would remain in balance when various pairs of clay balls were placed on it was sufficient in itself to produce conservation of substance for clay figures. This observation supported the suspicion that had been voiced by Braine (1962) and others that children might actually have mastered the intellectual operations required for conservation at earlier ages but failed to demonstrate conservation because they lacked the clearly defined dimensional concepts that were required. If this proved to be true, then it might be that the task of speeding up the child's mastery of operational thinking would be much easier than Piaget's theory would indicate--that it would not be necessary to speed up the acquisition of very fundamental logical operations but merely to teach children dimensional concepts so that they could make better use of the operations they already possessed.

Attempts to replicate the "balance" finding with children from average to disadvantaged backgrounds failed utterly, however. The original experimental design called for training 20 non-conserving kindergarten children to a criterion of 6 successive correct anticipations on balancing problems and comparing their performance on a conservation posttest with yoked control subjects who were retested after the same amount of elapsed time. The experiment was terminated after the running of 12 experimental and 12 control subjects, however, when not one subject in either group showed any indication of improvement on posttest performance, even though the experimental children did succeed in mastering the balance problems.

The subjects in this study seemed to be totally immune to cognitive conflict or incongruity. Although all the experimental subjects eventually learned the simple rules for predicting which way a balance would tip when unequal balls were placed on it, they never seemed particularly perturbed if their predictions proved wrong. After the posttests I would attempt various maneuvers with some subjects to get them to see the contradictions inherent in their judgments. Having, for instance, elicited the judgment that the piece of clay shaped like a triangle contained more clay than the ball, I would then remove bits of clay from the triangle one at a time until the child judged that it now contained the same amount of clay as the ball. I would then mold the triangle into a ball, which the child could immediately see was smaller than the first ball. We would then go through the process of putting bits of clay back on until the smaller ball was the same size as the original one. This, of course, resulted in adding back all of the clay that had been taken off before, a fact which was pointed out to the child. One of the balls was then molded into another shape and the process repeated. The children could apparently go through such cycles indefinitely, with only mild interest, and without ever seeming to find it remarkable or puzzling that the amount of clay taken away to make the dissimilar objects alike in quantity was exactly the same as the amount that had to be replaced to make them alike when they were both balls.

The most obvious difficulty was that the children did not seem to be particulary interested in amounts of clay. When asked the question, "Does this one have more clay or less clay or the same amount of clay as this one?" the children often gave the impression of merely being devoid of any opinion on the matter. If many of them seemed not to understand the question, it was also the case that they seemed not to try very hard to understand it.

David W. Brison, who had been working as a research assistant on the conservation project, devised a new attack on the teaching of conservation through the use of a social situation in which comparisons of quantity and the use of past information in quantitative judgments would assume importance to the children. The following paper reports a test of this approach, which Mr. Brison conducted as his doctoral dissertation research.

2.52 Problem and background

Experimental induction of the concept of conservation of substance was the primary concern of this experiment. Conservation of substance is attained when the child notes that the total amount of a substance remains constant even though shape or geometric form is altered. One method of testing for the presence of this concept is to present two clay balls of the same weight and diameter. After emphasizing the equality of substance, the E deforms one of the balls into another shape (cross, sausage). The child is then asked whether the deformed object has less than, more than, or the same amount of clay as the ball. Liquid substances can be used by utilizing various shaped containers.

Piaget and his colleagues have extensively investigated the concept because of their belief that important cognitive operations, namely decentering and reversibility, are involved in the conservation task. Decentering, as seen in the conservation task, is the ability to simultaneously account for perceived change in two dimensions. The nonconserving child attends to change in only one dimension such as length or width and therefore asserts that the deformed object has either more or less clay. The conserving child realizes that change in one dimension (height) is compensated for by change in another dimension (width). Reversibility, in the conservation task, is the ability to see that the deformed object can be transformed back into its original shape. These operations enable the child to extend his interaction with his environment because he is no longer tied to direct physical action in his representational thought but is able to rely on mental reversibility of his actions.

Piaget has not attempted to accelerate the acquisition of the concept but has studied its manifestations in both conserving and non-conserving children. Other investigators (Bruner, 1961; Smedslund, 1961a, 1961b, 1961c, 1961d, 1961e, 1961f, 1962) have tried to induce conservation of substance experimentally. Smedslund has tested hypotheses derived from both external reinforcement learning theory and equilibrium conflict explanations of cognitive development. It is generally agreed that Piaget has said relatively little about these rules of transition (Flavel, 1963; Hunt, 1963; Kessen, 1962) one exception being his treatment of accommodation which is the change in intellectual structure necessitated by demands which the environment makes on an individual. Studies of attempts to accelerate conservation of substance have been reviewed in Brison's unpublished doctoral dissertation (1965), in Flavell's book (1963), and by Wallach (1963). These reviewers agree that previous attempts to induce conservation of substance experimentally have met with only modest and tenuous success and this has impeded efforts to test hypotheses derived from various theories.

The present study was different from most other attempts to induce conservation, such as Smedslund's, in that acceleration itself was the primary purpose not the testing of hypotheses derived from opposing theories. This strategy seemed justified because of the previous resistance of the concept to experimental induction. If the acquisition of conservation can be accelerated under controlled experimental conditions in a relatively short period of time, then subsequent research can isolate and test theoretical explanations of the learning which occurred.

There are several possible reasons why previous attempts to induce conservation have not been successful.

The children used in these experiments might not have understood the words more, same, and less which are used in the conservation test. Furthermore, there is the possibility that the children might not have understood the word amount and might have equated it with length or height or applied it inconsistently to whatever aspect of the stimulus commanded their attention. An additional explanation is that the experimental conditions may not have been motivating or that attention to the training was not adequately maintained.

In this experiment the children were pretested for the understanding of more, same, and less. An attempt was also made to get the children to apply these words unambiguously to the conservation situation by asking them whether the glass the liquid was poured into had more (same, less) juice to drink. It is then difficult to apply the terms to anything other than amount. In regard to the problems of motivation and attention, the attainment of a larger reward was contingent upon acquisition of the concept of conservation. This feature was introduced not only to heighten general motivation to succeed but also to produce in the child definite expectations, the reversal of which appears to have motivational properties (Charlesworth, 1964; Hunt, 1961).

This experiment was also designed so that properties of the experimentally induced concept could be studied, specifically: (1) transfer of conservation to materials not used in experimental training, (2) stability of the concept when the child was faced with an apparent contradiction, (3) relevance of a set to seek explanations and search for causes.

2.53 Method

The subjects were pretested on verbal concepts (more, same, less), conservation of substance, and explanatory set. Those who showed no evidence of conservation of substance were divided into an experimental group (N=24) which received training and a control group (N=26) that received a placebo treatment. The control and experimental groups were matched on their explanatory set scores. Twelve subjects who clearly demonstrated conservation of substance on the pretest were used as accessories in the experimental training. All subjects were posttested on conservation of substance and those who gave acceptable conservation responses were administered an extinction item. The temporal order of events was: 1st day, pretests; 2nd and 3rd days, experimental training (25 min. per day); 7th day; posttests and extinction.

Subjects

The subjects were 62 kindergarten children from a middle-class suburban community. They ranged in age from 5 years, 4 months to 6 years, 4 months with a mean of 5 years, 7 months.

Apparatus

Materials employed for the conservation of substance questions were clay in 3 1/2 oz. balls, white sand, and a red fruit drink. The drink was also used in the experimental training and placebo sessions. Regular cylindrical drinking glasses served as containers. The extinction item required a glass with a thick bottom, but the other glasses did not have this type of bottom.

Procedure

Pretests of Verbal Concepts. The examiner showed the subject two identical cylindrical containers with unequal amounts of juice and asked successively which glass had less juice to drink, more juice to drink, and did they have the same amount of juice to drink. After equalizing the amounts, the S was asked if they had the same amount of juice to drink.

Conservation of Substance Pretests. The conservation of substance tests used the familiar format originated by Piaget and used by others. Conservation questions were given only to subjects who answered all four verbal pretest questions correctly. The general format was as follows: the examiner presented equal amounts of substance and stressed the sameness of amount. Then one of the objects was made into another shape. The experimenter asked if the new object had more, the same amount, or less than the other object. The subject's reaction to this question was called the prediction. The subject was then asked to explain his prediction. Simple restatement of the prediction drew the response, "Why do you think they are the same (more, less)?"

The order of pretest conservation items was as follows:

1. Two identical cylindrical glasses of juice: juice of one was poured into a narrower glass (on all juice items, the experimenter asked if the new glass had more, same or less juice to drink than the original glass). This item was administered by the experimenter immediately after the verbal pretests. This was a screening question done individually in a corner of the kindergarten class. The children were then given the rest of the conservation questions and the explanation questions by another examiner in another room.

2. Two orange balls -- one changed into a ring.

3. Two green balls -- one changed into a triangle.

4. Two identical cylindrical containers with equal amounts of sand -- poured into wider glass.

5. Two identical cylindrical containers with equal amounts of sand -- sand from one poured into narrow glass.

Conservation items 2,3,4,5 were not administered by the same experimenter as item 1.

Conservation Scoring Categories. All explanations irrespective of the accuracy of the predictions were scored either acceptable (Acc) or unacceptable (U). Acceptable explanations represented some attempt to organize the features of the situation into an explanation. Perceptual responses (it looks bigger) were scored Acc. An unacceptable explanation was one that did not represent an attempt to formulate an explanation. Common examples were insistence on restating the prediction and referral to an outside source (i.e.,"my mother told me so").

Explanations of all correct predictions were scored as:

1. Adequate --- two categories of adequate responses were scored: (a) compensation -- any reference to the fact that change in one dimension is compensated by change in another dimension, (b) reversibility -- referral to the original equality of the amounts. The explanation that nothing had been added or taken away from the object before it was deformed was included in this category.

2. Inadequate -- explanations which were obviously wrong.

Three hundred fifteen explanations were scored as acceptable or unacceptable by two independent judges. The percentage of agreement was 96 and a phi correlation coefficient of .84 was obtained. In 108 explanations of correct predictions scored as adequate (compensation or reversibility) or inadequate, the judges agreed on 94 percent of the classifications.

Explanatory Set. The explanatory questions were included in the pretesting in order to see if the general tendency to look for causes and formulate explanations had any relevance to the experimental induction of conservation in nonconserving children. The experimenter administered the explanation questions upon finishing the conservation pretesting. The questions were:

1. Why do houses have doors?
2. Why do books have covers?
3. Why can older boys (girls) usually run faster than you can?
4. Why do people read?
5. Why is it harder to ride a bicycle than a tricycle?
6. Why don't people spank little tiny babies?

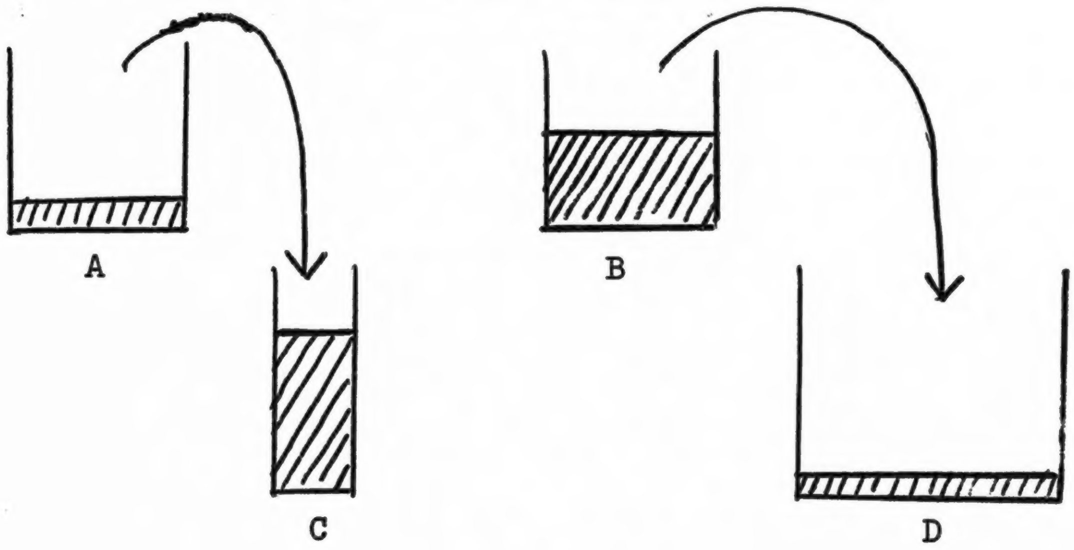
The subject's explanatory set score was the total points he received on the explanation questions plus one point for each acceptable explanation on conservation questions.

Criteria for Inclusion in Experimental and Control Groups. The subjects had to pass all four verbal concept questions and fail the first conservation of substance (juice) question by giving a wrong prediction. On the four additional conservation questions the subjects did not give any correct predictions with adequate explanations and not more than two correct predictions with inadequate explanations. Each group (experimental and control) gave a total of only four correct predictions with inadequate explanations. The subjects were matched on their explanatory set scores and randomly assigned to both groups.

Criteria for Selection of the Conservation Group. Ten of the twelve subjects in the conservation group gave at least four correct predictions with adequate explanations in response to the pretest conservation questions. One subject predicted correctly four times with three adequate explanations. The remaining child had two correct predictions with adequate explanations.

Experimental Training: First Day. The experimental groups were divided into six subgroups and two conserving subjects assigned to each subgroup. The subjects in each subgroup were shown two identical cylindrical glasses A and B (Figure 2.53.1) with obviously unequal amounts of juice. Figure 2.53.1 represents the process followed. A was poured into C and B into D. The examiner stated, "I know that you all like juice, which one (C or D) would you take if you wanted more juice to drink?" The children pointed to the glass they wanted. The liquid was then poured back into glasses A and B and the Ss given the amount they chose to drink. A child who chose the correct glass was asked to give an explanation. This procedure was repeated two more times, using different pairs of unequal glasses. On the first trial, almost without exception, non-conserving subjects chose the narrow glass and conserving subjects the wide glass.

Figure 2.53.1 Glasses Used in Experimental Training.



Arrows indicate pouring of juice in the first days experimental training.

Experimental Training: Second Day. The second session involved a variation to prevent the subjects from concluding that the wide jar always contained the larger amount. Each child received two trials (conducted again in group setting); on one trial the largest amount was poured into the wide glass and on the other it was poured into the narrow glass. Subjects answering correctly were asked to give explanations.

Control Group Placebo Treatment. The control group was also divided into six subgroups and participated in placebo treatment sessions according to the same schedule as the experimental group, except that the main activity was free construction using lots of material similar to those used in the experiment described in section 2.4. At intervals during the sessions juice was served -- the same number of times as in experimental sessions -- but equal amounts were measured out for all children.

Conservation Posttest. The posttest was identical to the pretest except that the orange balls were changed into a cup and the green balls into a cross.

Extinction Items. These questions were administered to all experimental, control, and conserving subjects who gave correct predictions to posttest items 1 (juice-narrow) and 5 (sand-narrow). Following a correct prediction on these questions, the experimenter returned the glass he had just emptied to his lap and replaced it on the table with a glass which was similar in all respects to the one removed except for a thicker glass bottom. The juice in the narrow glass was poured back into the "fake" glass and the discrepancy in levels pointed out to the subject. The subject was given a chance to respond and then asked to explain the discrepancy.

Examiners. Five experimenters assisted the principal investigator and in no instance did a posttest examiner know whether the subject he was testing was in the experimental or control group.

2.54 Results

Induction of Conservation of Substance. Table 2.54.1 shows the main results of the experimental induction. Twelve of the 24 experimental subjects who received training showed some evidence of conservation on the posttest as opposed to only one of the control subjects. The criterion for evidence of conservation was at least one correct prediction with an adequate explanation. The 12 experimentally trained subjects who evidenced conservation on the posttest gave a total of 38 "same" responses; 3.2 of these responses were accompanied by an adequate explanation. The minimal criterion of one same response was exceeded in all but one case. A significant difference between the experimental and control subjects was found by chi-square analysis (χ^2=11.52, $p < .01$).

Table 2.54.1 - Evidence of conservation

Group	Yes	No
Experimental (N = 24)	12	12
Control (N = 26)	1	25

Table 2.54.2 presents the number of experimentally trained subjects who reached a criterion of four correct predictions on the posttest. A chi-square analysis reveals that the difference between experimental and control groups is significant (χ^2= 3.93, $p < .05$).

Table 2.54.2. - Experimental Ss exhibiting four "Same" predictions

Group	Yes	No
Experimental (N = 24)	5	19
Control (N = 26)	0	26

The experimental group gave a total of 41 correct conservation predictions in the posttest and the control group a total of 3. A nonparametric median test showed this difference to be significant at the .001 level.

Transfer to Substances Not Used in Experimental Training. The criterion for acquisition of conservation in a specific substance (clay, sand, juice) was at least one correct prediction with an adequate explanation. Of the 12 subjects who demonstrated conservation on one or more items, 6 demonstrated it on juice, 8 on sand, and 9 on clay. Thus conservation responses were most often evidenced on clay, although clay was not used in training and does not share with the training substance (juice) the property of being pourable.

Transfer of conservation from one substance to another for the experimentally trained subjects is shown in Table 2.54.3. The transfer between clay and sand is significant (Fisher exact test, $p < .025$) but not between juice and the other substances. The number of questions differed (one for juice, two each for sand and clay) and this complicates the analysis.

Responses to Extinction Questions. Table 2.54.4 summarizes the extinction data. A response was defined as relinquished only if the subject stated that the glass with the thick bottom contained more of the substance. Every other response was defined as a resisting response. Three experimental subjects

Table 2.54.3. - Transfer of conservation for experimental Ss, as shown by pass-fail contingencies, with one correct prediction with explanation as the criterion for passing

		Clay	
		Pass	Fail
Sand	Pass	6	2
	Fail	3	13

		Sand	
		Pass	Fail
Juice	Pass	3	3
	Fail	5	13

		Clay	
		Pass	Fail
Juice	Pass	4	2
	Fail	5	13

Table 2.54.4. - Responses to extension items by experimental conservers and conservation group

	Experimental	Conservation
Resisted on both questions	2	6
Relinquished on both questions	0	2
Relinquished on 1st, resisted on 2nd	1	3
Took 1st item (relinquished or resisted) and gave wrong prediction on last conservation question	5	1
No more same predictions after 1st extinction question	1	0
Only took 2nd extinction question		
relinquished	2	0
resisted	1	0

responded with a wrong prediction to the first juice question and were not administered for the first extinction item. Six experimental subjects gave a wrong prediction on the last conservation question after taking the first extinction item. Five of these subjects did give correct predictions to other questions indicating that the concept was not totally extinguished.

Only five of the twelve experimental Ss who showed evidence of conservation reached a criterion comparable to the level of the Conservation Group. The performance of these five Ss on extinction was:

1. two resisted twice.
2. one relinquished on the first, resisted on the second.
3. one took only the second and resisted.
4. one took only the second and relinquished.

Although not statistically significant because of the small number, this performance resembles the conservation group's reaction to extinction.

Relevance of Explanatory Set. The explanatory set score was not related to acquisition of conservation for the experimentally trained subjects. The difference between means of acquiring versus nonacquiring experimental subjects on the explanation questions approached significance ($p < .10$, one-tailed test).

2.55 Discussion and conclusions

The data show that the experimental training was influential in accelerating the acquisition of conservation of substance. Five experimental subjects gave at least four correct predictions while seven other experimental subjects showed significant growth in the development of the concept.

The experimentally induced concept was not restricted to the substance used in training. This transfer phenomenon in respect to clay is even more remarkable in view of the finding by Hyde reported by Lovell (1965) that some children who conserved with liquid failed to conserve with clay.

The stability of the experimentally induced concept, measured by the child's reaction when faced with an apparent contradiction of the conservation principle appears similar to the stability evidenced by the conservation group who had the concept before the experiment started. This statement is based on the performance of the five experimental subjects who demonstrated conservation at a level commensurate with the conservation group. Smedslund (1961c) applied a somewhat different extinction item to eleven subjects who acquired conservation by external reinforcement and all of them extinguished.

The data from induction, transfer, and extinction indicated that the experimentally induced concept is like the concept which children acquire "naturally."

The main limitation of this study is the failure to isolate the variables in the experimental training responsible for the induction of conservation. The experimental training was standardized and, in light of the difficulty other investigators have had in producing the concept, it took place over a relatively short period of time. Subsequent research will have to manipulate the observed variables operating in the training.

One criticism of attempts to accelerate conservation has been that the experimental training provides practice in the use of the "same" response, which children are sometimes reluctant to use. It is important to note that nothing in the experimental training would contribute to the increase in same predictions if the concept of conservation of substance was not acquired. Training was in the conservation of inequalities in which the correct (reinforced) response was always more or less, and not the same.

In the conservation training, the subjects had the expectation that the narrow container had more juice and this expectation was disconfirmed. Then in order to obtain more juice they had to integrate the elements of the conservation situation. A logical explanation was offered in the presence of conflicting perceptual cues. This condition was motivating and resulted in concentrated attention.

In Piagetian terms, the child was maintained in a condition of socialization -- a situation which required coordination of other viewpoints. The most important aspect of the training was the demands made on the child by the training situation. The subject's existing cognitive structure could not adequately cope with these demands and therefore these structures were in a state of disequilibrium.

All the experimental Ss evidently acquired some mastery of reversibility during the training sessions, in that they were able to determine correctly which of two dissimilar glasses contained the more juice, when the only reliable information was that provided by comparison of the two quantities in their original, identical containers. In this respect, all the conservation tasks represented transfer problems; and so the fact that half of the experimental Ss gave conservation responses on the posttest is evidence of considerable transfer of the reversibility operation to new and different problems.

2.56 Summary

Twenty-four nonconserving subjects received experimental training designed to induce conservation of substance. Twenty-six matched control subjects were not trained. Training was in the conservation of inequalities of liquid in a situation where the subject's expectation of an event was reversed. The child had to integrate the elements of the conservation situation to obtain a desired reward. Twelve of the experimental subjects showed evidence of acquiring conservation. Five of these subjects gave at least four of five correct conservation predictions. Conservation transferred to substances (clay, sand) not used in experimental training. On extinction items the five experimental subjects with four correct predictions performed similarly to subjects possessing conservation before the experiment. The theoretical implications of the results were briefly discussed.

2.6 Teaching Formal Operations

Siegfried Engelmann

2.61 Prefatory note (by Carl Bereiter)

Of the few educational implications that have been drawn from Piaget's work, the most obvious and most widely endorsed is that the instruction of pre-adolescent children must be founded upon concrete operations (Ausubel, 1963; Flavell, 1963; Bruner, 1961). The full weight of Piaget's developmental theory seems to lie behind this implication. For in Piaget's theory the logic of childhood is not general but has concrete reality (past, present, and imminent) as its domain of application. The child's thinking is entirely geared to dealing with the actual, not out of choice or aversion to the abstract, but because his cognitive structures are only capable of dealing with empirical givens and cannot generate hypothetical instances. It is only as the child's structure of concrete operations becomes over-elaborate and unyield that he begins to develop second-order operations that allow him to manipulate rules rather than facts, in ways that will lead to conclusions that are valid for the full range of possibilities to which the rules apply (Piaget, 1960).

Although the theoretical point that Piaget makes about the fundamental difference between the concrete mode of logical operation and the formal or abstract is clear enough, the behavioral referents of this distinction are not nearly so clear, and for this reason the implications for educational practice are vague. Let us first rule out a few possibilities.

1. To claim that the thinking of preadolescent children binds them to concrete reality cannot mean that they are incapable of dealing with imaginary or hypothetical instances--or if it does the claim is patently false. From Mother Goose days on, the child is expected to apply to imaginary instances the same reasoning operations he applies to actual ones. Piaget himself, in studying the development of moral judgment in children, used imaginary instances as a vehicle for exploring the thinking of children as young as seven (Piaget, 1932). Convincing normative evidence is supplied by the Stanford-Binet Intelligence Test, on which half of eight-year-olds are found capable of dealing with problems such as detecting the absurdity in the story of an old man who used to walk around a park every day, but now, because he tires more easily, walks only half the way around and then back. Such a problem seems to require a highly sophisticated ability to deal with the hypothetical.

2. The young child's correct use of semantic rules implies that he is capable of applying "rules about rules" rather than merely rules about facts. The following interchange took place between parents and a daughter, age five.

MOTHER: It's warm out today.

FATHER: Not only that, but it's not cold.

DAUGHTER: Silly, warm and not cold are the same thing.

The child's generalization obviously applies not simply to the weather but to the entire domain of the terms warm and not cold. She is stating flatly that two propositions are equivalent. Although not every child would express the rule, most would behave according to it. No five-year-old, of course, would be expected to deliver the fully abstract rule that the negation of the negation of a proposition is equivalent to the original proposition, and yet it would appear that the child in the example was operating according to such a fully general rule--that she would treat all other polar attributes that she understood in the same way as she treated warm and cold.

3. Concrete operations do not require concrete objects. Ask a child how many times he fell down while skating. The question causes a child no bewilderment in spite of the fact that there are no objects to count. Unless the child has been poorly taught, he conceives of arithmetic facts and rules as being true in general and he is comfortable in operating upon numbers rather than real objects. In one experimental kindergarten for disadvantaged children, the concept of _pounds of force_ was introduced as an equivalent of "how hard you could push." The children were immediately capable of applying arithmetic to the solution of numerical problems involving the resolution of opposing forces. Here there was nothing at all to be counted. The children were, in fact, treating the real numbers as a lattice model which could be fitted formally to the problem in hand--the very hallmark, according to Piaget, of formal logical operation.

The above observations all indicate some kind of ability to operate with the hypothetical and the general at ages well below adolescence--at ages, in fact, below those at which _concrete_ operations are supposed to be developed! What, then, is peculiar to the later years? There is, first of all, the elaboration of formal logical thinking into a complete system that characterizes the mature person's thinking generally rather than only particular aspects of it. This characterization is seldom fully general, however, as Piaget has noted (Piaget, 1960). In all but the most sophisticated theoretical journals it is deemed necessary to buttress every formal argument with an empirical example in order to ensure comprehension and conviction. A more marked difference claimed by Piaget is that in later years the child becomes able to generate the logical possibilities that are inherent in problems. Thus the child is no longer bound to dealing with possibilities that are directly presented to him or generated out of concrete operations, but is able to make systematic use of the hypothetico-deductive method. These appear to be the major claims made by Piaget as to the distinctiveness of the level of formal logical operations--its integrated structure, for which Piaget has written a logic, and its hypothesis-generating capabilities.

But neither of these claims carries the least implication that there is anything wrong with teaching abstract material to younger children. Indeed, it would seem to be a very good idea for developmental reasons, not to mention practical educational ones. If the child is eventually to integrate formal logical operations into a complete system, it would seem to be a good idea to provide him with some prior, if limited, experience in abstract thinking, and some abstract principles to integrate. And the fact that the younger child may not be able to generate logical possibilities, instead of suggesting that education for such a child must stick to the concrete, might be taken to indicate that teachers should supply the logical possibilities for children to work with or give them explicit direction in generating them.

An insistence on teaching through concrete operations can be severely limiting in the teaching of mathematics and science. It forces the teacher to accept a baggage of gimmickery so heavy that it is hard to travel very far in a school year. Compare the bizarre devices and the length of time devoted to giving children the "feel" of negative numbers, to the simple formal statement that a negative number is a number less than zero. Insistence on concrete operations makes the teaching of social science, as science, virtually impossible. But where concreteness becomes absurd is in the teaching of thinking or problem-solving strategies. For any practice that deserves the name strategy, rather than recipe, must apply to problems that are structurally the same but different in concrete particulars--hence, it must be a formal operation. It may (conceivably) be learned through work on a succession of concrete problems, but the concrete problems would have to be chosen because of their formal properties. Thus the teaching of strategies, no matter how concretely it may be presented, must be directed toward the teaching of formal operations--operations that transcend any particular set of real or imagined physical operations.

The following experiment, conceived and carried out by Siegfried Engelmann, explored the possibilities of teaching a formal method of handling a class of problems that differed radically in their concrete aspects--which could not be solved by any uniform concrete operation--and teaching it to children who by age and performance were below the level even of concrete operations. Concessions to the immaturity of the learners are evident first in the fact that the logical possibilities are taught to the children if they are unable to generate them and secondly in the fact that the possibilities are limited in number--two, in each problem. Neither of these concessions, however, detracts from the abstractness of the operation being taught.

2.62 Nature of the study

The purpose of this paper is to report on an experiment in which culturally privileged and culturally deprived preschool children were taught the rules for handling a class of logically similar formal operations. The hypothesis investigated was that although these children were not yet at the concrete operational stage (as evidenced by their inability to exhibit conservation of liquid quantity) they could learn to solve problems that require the use of hypothetical-deductive reasoning.

The type of formal operations presented in the present experiment is based on a relatively precise definition offered by Inhelder & Piaget (1958, p. 279). They suggest that one can infer whether the subject is using formal operations from the proofs employed by the subject.

> If they do not go beyond observation of empirical correspondence, they can be fully explained in terms of concrete operations, and nothing would warrant our assuming that more complex thought mechanisms are operating. If, on the other hand, the subject interprets a given correspondence as the result of any one of several possible combinations, and this leads him to verify his hypotheses by observing their consequences, we know that propositional operations are involved.

Problems that require the latter type of solution can be created. The simplest design would be a problem in which there is no real "empirical correspondence" but only possible correspondence; in other words a problem posed in hypothetical terms. The problem would be designed so that all but one of the possible solutions of the problem are eliminated through the use of propositional reasoning.

Piagetian tasks were not selected for the present experiment because they do not represent the least complicated expression of the theoretical principles involved and because they are usually based on exceptional, empirically limited principles. For purposes of instruction, it was necessary to use a broad set of problems all based on the same fundamental principle so that all could be taught in a similar manner. These considerations necessitated the development of a set of problems which can be referred to as the "center problems." Whether or not these are actually formal problems can certainly be argued, according to the definition of formal operations one accepts. According to one interpretation, they are formal because they require the formation and manipulation of propositions-about-propositions. According to another, they are not because they do not occur as part of a structured whole.

The center problems were so named because in each there is a logical or mechanical relation between two and only two members. The problems in the set are analogous and require analogous solutions, which can be demonstrated by noting the similarity in structure of the following problems, and concomitant similarity in the information-organizing processes that are required to solve them. Three problems are compared. Please note that the propositions provided are not necessarily exhaustive or even sufficient for the solution of each problem. They simply represent some of the propositions that are necessary to solve the problem.

1. Piston problem in which a single piston is connected to a single crank. The problem involves determining how the system became activated.

Propositions.

a. If the piston goes down, the crank turns.

b. You can make the piston go down either by setting off the spark plug or turning the crank.

c. If you set off the spark plug, a loud exhaust noise will be heard.

d. If a loud exhaust noise is heard, you set off the spark plug.

e. A loud exhaust noise is heard.

f. Therefore, you set off the spark plug.

2. A freshly painted floor, on which tracks are made by a rolling object which careens against the wall and is "reflected" making a V-shaped

path. The problem involves determining whether a given path was made by rolling an object from side A or side B. A person can get to side A only by walking through the paint.

Propositions

a. If path A (the left member of the V path) becomes more nearly parallel to the wall, path B becomes more nearly parallel to the wall.

b. You can make path A more nearly parallel either by rolling the ball from a new point on side A or by rolling the ball from a new point on side B.

c. If you roll the ball from a new point on side A you will get paint on your feet.

d. If you get paint on your feet, you rolled the ball from a new point on side A.

e. You have paint on your feet.

f. Therefore, you rolled the ball from a new point on side A.

3. The double-H model (an H with 2 parallel horizontal bars), which is of particular interest because it is a simple abstract version of a center problem. Certainly it is not the only possible simplified form, but it is a legitimate one because it generates the same set of propositions as the other center problems, and yet it is stripped of all but the properties of sameness, longer, shorter, etc.

Propositions

a. If A (the top horizontal bar) is shorter, B (the bottom horizontal bar) is longer.

b. You can make A shorter either by shortening A or by lengthening B.

c. If you shorten A, A will not reach the vertical line. If you lengthen B, B will stick out past the vertical line.

d. If A does not reach the vertical line, you have shortened A.

e. A does not reach the vertical line.

f. Therefore, you have shortened A.

This problem appears to be one that can be solved on a simple, "visual" level; however, it cannot be. That such a sinuous chain of reasoning is actually necessary can be demonstrated quite easily.

There is no way one can determine how two apparently identical double-H models differ unless one understands the function of the vertical lines. These must be understood not as mere visual elements but as rule-type statements. Did they move or did they remain stationary? Measuring the difference or trying to judge the difference visually begs the question by helping one form the necessary proposition. Also, it must be kept in mind that any argument used to simplify the solution of the double-H problem will also perforce apply to the solution of all other center problems, since the various center problems are analogous.

2.63 Method

Subjects

Subjects were 10 preschool children (ages 3 1/2 to 5 3/4) half of whom were relatively culturally deprived Negroes, the other half of whom were relatively privileged Caucasians. Subjects were selected on the basis of teacher judgments that they were the "best talkers" in their respective nursery schools in Champaign--one a summer nursery school for Negro children, and the other a relatively high prestige year-round nursery school. The two groups of children were designated "culturally deprived" and "culturally privileged."

Training Procedures

No pretest was administered to determine whether the children would be able to solve the criterion problem. It was felt that the presentation of the problem (since it is analogous to those the children were to study) would merely serve to direct the children's efforts and contaminate the results.

Daily training periods were held with each group during July and August, 1964. These lasted about 20 minutes, 5 times a week, for 5-7 weeks. No firm termination procedure was adopted; rather, the criterion problem was presented when the author thought that the children were ready to solve it. The training sessions were aimed at teaching the information--the rules--necessary for the solution of specific center problems. The children were taught, step by step, for instance, how to solve the problem in which a bar (half of which was covered with fresh paint) was either pushed or pulled through a hole in a wall. They were taught the two ways in which a given change can be achieved in the system (either by pushing on one side or pulling on the other). In another series of exercises, they were taught how to use such rules as, "If you touch the jam (or wet paint), you'll get jam (or wet paint) on your hands." They were then taught what to do with this rule when they wanted to "go backwards" and determine what had happened. "Turn the rule around; if you've got jam on your hands, it means you've been in the jam." They were also taught the basic principle of change that was to be used in the criterion problem (a balance-board problem). "You can tilt a balance board either by pushing up on this side (A) or down on this side (B)." What they were not taught was how to combine the rule about the paint with the principle of the balance board in a chain of deductions that would lead to the proper conclusion. The purpose of the experiment was to determine whether or not transfer of the basic reasoning pattern for handling center problems would occur. The Ss were provided with all of the specific information needed to achieve the transfer. However, in transferring a reasoning pattern, it is necessary to modify it. This modification was not taught.

Instruction began with the root concepts, same-different, long-short, etc., and proceeded through basic propositions involving the concepts and and or, to the more complicated deductive chains.

Criterion Problem. The children were tested on a problem involving a tilted balance board. To ensure the use of "propositional reasoning" the problem was presented in two steps. To pass the children had to indicate an awareness of both the possible ways in which a given change may be achieved in the system. (They had been taught this principle.) Then they had to eliminate one of the possibilities by reference to a familiar limiting condition (the point). The application of this condition to the balance board had not been taught.

A diagram of a tilted balance board was presented to the children individually. They were asked, "What could somebody have done to make the board look this way?" The answer: "Push down on the left side or push up on the right side." A heavy line of red chalk was then drawn across the top of the balance board. Its function was explained. "This is fresh red paint. Now think big and see if you can figure out what happened if I tell you that the guy who moved this board got red paint on his hands."

2.64 Results

Criterion Problem.

The performance of the subjects in working the criterion problem is summarized in Table 2.64.1. Four of the five members in the culturally privileged group successfully solved the problem. Three out of five in the culturally deprived group solved it. All of the five-year-old children in the experiment solved the problem. The only non-five-year-old child who passed the criterion problem was a four-year-old in the culturally privileged group (who also passed the test of conservation of liquid quantity.

Table 2.64.1. - Performance on criterion problem and conservation test

Culturally Privileged					Culturally Deprived				
Name	Sex	Age	Criterion Problem	Conservation Test	Name	Sex	Age	Criterion Problem	Conservation Test
Au	F	5	Pass	Fail	Ra	F	5	Pass	Fail
Ly	F	5	Pass	Fail	Er	M	5	Pass	Fail
El	F	5	Pass	Fail	Ju	M	5	Pass	Fail
Da	M	4	Pass	Pass	Sh	F	4	Fail	Fail
Er	M	3	Fail	Fail	De	F	4	Fail	Fail

The results of the criterion test do not fully show the differences between the groups. The two children in the deprived group who did not pass the criterion problem received a full seven weeks of training. Four members of the privileged group (Au, Da, Ly, and Er) received five weeks of training and learned, in addition to the center problems, how to work relative direction problems (A is west of B but east of C), mirror problems, and sophisticated three-step non-center deductions! During the same period of time the two children in the culturally deprived group were just beginning to learn such fundamental notions as "if this end of the board is down, it is not up." They were only starting to learn the basic language assumptions. It took over a week to teach them to say the word or.

Test of Conservation

After the criterion problem and been administered, the children were given a test of conservation of liquid quantity according to the same test procedure as described in section 2.53. The conservation test was given to provide an independent Piagetian indication of the cognitive maturity of the children. If they passed the criterion problem (indicating the ability to handle formal-type reasoning) and at the same time failed the test of conservation (placing them below the level of concrete operational thought), the Piagetian interpretation of growth would be caught in something of a dramatic paradox. One of the most important issues raised by such a paradox would be concerned with the role of specific information in determining a child's performance on such tasks as conserving liquid as well as handling logical operation. Perhaps the only reason most five-year-old children do not conserve is because they do not understand the properties of water, namely the property of being non-elastic even though it looks elastic when it seems to blossom from the end of a sprinkler or when it assumes the shape of the container into which it is poured. There are logically as many different types of inability to conserve as there are constituent concepts and skills in the conservation task, which means the entire process can be conceived in terms of specific information. As indicated in Table 2.64.1, only one child passed the conservation test.

2.65 Discussion

Since only one of the seven children who passed the formal problem exhibited the ability to conserve liquid quantity, it is evident that children who have not yet achieved the level of concrete logical operations can be taught to handle problems of a formal logical character. Although the present experiment did not deal with a large number of subjects, large numbers are not necessary to discredit universal statements. The statement, "Steps 1 and 2 are necessary to get to level 3" can be thoroughly discredited if one example can be cited in which level 3 was achieved without taking the steps 1 and 2. In the present experiment, most of the subjects achieved what amounts to level 3 without taking the other steps. It follows, therefore, that the occurrence of these steps under normal conditions tells little about what can be achieved under conditions of purposeful teaching.

2.66 Summary

Five culturally deprived and five culturally advantaged preschool children were systematically taught the skills, concepts and the basic argument form necessary to handle an analogy class of what has been rather loosely defined here as "formal operational problems." After the 5-7 week training period, the children were tested on a criterion problem to determine whether or not the training would transfer (thus indicating ability to handle the operation). Children also received a test of conservation of liquid to provide a comparative measure of cognitive maturity. The basic hypothesis was that through the training program, children who were at the preoperational stage as measured by the test of conservation of liquid quantity would be able to perform at the formal operational level (cr a more abstract level of cognitive performance) on the criterion problem. Thus, the order of cognitive development suggested by Piaget would be seriously violated.

The hypothesis was confirmed, whether one assumes a strict interpretation of formal operations and therefore maintains that the criterion problem was not formal, or whether one accepts the propositions-about-propositions interpretation of formal operations. According to either interpretation, the experiment shows that children can, with instructions, operate at a cognitive level well above that at which they function normally. All of the five-year-old subjects (three culturally deprived and three culturally privileged) successfully solved the criterion problem. None of these children passed the test of conservation of liquid quantity. Furthermore, the result was achieved through direct instruction, thus prompting the question: Is it possible to "teach" any formal operation directly through the presentation of specific information?

While the present experiment does not answer this question, it shows that when one is dealing with a fairly definite set of formal operations which are similar or analogous in certain gross features, one can teach very young children how to use them. There is no very obvious reason why this procedure would not produce similar results with any set of logical operations. There is no reason to abandon the hypothesis that a child learns what he has been taught.

2.7 Resume of Exploratory Studies

The studies reported in sections 2.1 through 2.6 were intended to explore various approaches to early education in several different areas of learning, with the object of identifying directions for innovative development. In this section each study will be summarized and certain implications will be drawn from the aggregate of findings.

1. Teaching Reading to Two- and Three-Year-Old Children. Five two-year-old children and five three-year-old children of average to above average language aptitude participated in five months of reading activities. The two-year-olds engaged mainly in free play with a variety of teaching machines and other manipulative devices; the three-year-olds engaged in teacher-directed

group games designed to teach letter recognition, word recognition, and eventually some phonics. In both groups little visible learning occurred until an activity was introduced that amounted to conventional flash-card drill. Half of the subjects attained some stable sight vocabulary in this way, ranging from 10 to 50 words, and one child acquired some ability to decode novel words. Scores on the Illinois Test of Psycholoinguistic Abilities indicated no relative gains in language-related abilities; but there was a significant interaction between treatment condition and changes in verbal versus non-verbal scores, children in the group games condition showing more favorable results on verbal tests while children in the teaching machine condition showed more favorable results on non-verbal tests.

2. The Effect of Free-Time Use of a START Teaching Machine on Reading Ability in the Kindergarten. The main comparison in this study was between seven kindergarten classes in which a teaching machine was placed for free-time use and seven matching control classes which had no machines. The machine was of a type used in the previous study, which presented stories both auditorily and visually and which required the child to make occasional choices between printed words in order to keep the story going. On a recognition test consisting of words used in these choice-frames, every experimental class did better than its matched control class. The difference, using classes as the sampling unit, was significant at the .05 level.

3. Children's Preferences for High Versus Low Frequency Words. This was an ancillary study conducted to investigate the possibility that the customary procedure of basing beginning reading materials on

words of high frequency may result in the use of words that have lower interest value than less frequently occurring words. The study used a paired-comparison technique, with 20 pairs of nouns and 20 pairs of adjectives. Twenty kindergarten children were given randomly balanced pairings of the full set of words, with order balanced within pairs. Words were presented on separate cards and for each pair the subject was asked to choose the word card he wanted to keep. For both nouns and adjectives, children showed a significant preference for high rather than low-frequency words. The results thus tend to vindicate the practice of using high-frequency words, at least for beginning readers.

4. Four Approaches to Construction Activities in the Nursery School and Their Relation to Creative Problem Solving. Groups of nursery school children were given six half-hour sessions of construction activities, using a large assortment of construction materials. Following this were three test sessions in which the children were assigned three construction tasks to be carried out independently without help or suggestions. The tasks were to construct a boat, a rattle, and a bird house. Photographs of the products were given blind ratings on the degree to which the product represented fulfillment of the task assigned (not on originality or esthetic value). A total of   children were involved in four replications of four treatments: Independent Problem-Solving, in which daily tasks were assigned as in the test condition; Independent Construction, in which no tasks were assigned but the children were merely encouraged to make something different each day; Guided Problem-Solving, in which teacher and children worked out the solution of construction problems cooperatively; and Guided Construction, in which the

teacher directed the children through a predetermined series of steps in the construction of prescribed objects. The analysis of ratings provided for the separation of the two treatment factors, guidance versus independence and problem-solving versus construction. Only the first factor showed any significant effect, and this was in favor of guidance over independence. There was no interaction between treatment factors; thus Guided Problem-Solving, which was expected to have greater transfer value to the test situation, produced no better results than Guided Construction, which is popularly believed to discourage creativity.

5. Acceleration of Conservation of Substance. Twenty-four non-conserving subjects received training in the conservation of inequalities of liquid in a situation where the child had to apply the principle of reversibility in order to be successful in choosing the larger amount of fruit juice to drink. Twenty-six control subjects received a placebo treatment that involved the same amount of experience with the measuring out and drinking of fruit juice. Twelve of the experimental subjects and one of the control subjects showed evidence of acquiring conservation. Five of the experimental subjects but none of the control subjects gave at least four of five correct conservation predictions. Conservation transferred to substances (clay and sand) not used in training. On an extinction item the five experimental subjects with four correct predictions performed similarly to subjects who already possessed conservation before the experiment.

6. Teaching Formal Operations to Culturally Advantaged and Disadvantaged Preschool Children. Five culturally deprived and five privileged preschool children were taught to handle an analogy class of

"formal operations." All problems were variations on the logical form, "If X, then A or B'; if X', then A' or B (where X and X' are the inverse of each other, as are A and A', B and B')." After training subjects were tested on a problem of the same form which they had not previously encountered and also were tested on ability to conserve liquid quantity (which provided another measure of development). Seven passed the criterion problem. Of these, only one exhibited conservation of quantity. The experiment thus casts doubt on the generalizability of Piaget's conclusions about developmental progression from preoperational to concrete operational and thence to formal operational modes of thought.

Let us first consider what the above studies indicate about what can be taught to preschool children. The results do not so much suggest a radical revision upward or downward of estimates of children's learning capabilities as they do a shift in focus from the general to the specific. The studies on reading, for instance, do not tell us much about the general readiness of children from ages two through five for reading instruction. They do suggest, however, that if one considers only sight vocabulary, then a considerable amount of learning could probably be achieved even with the younger children by concentrating all of the teaching on this objective. The amount of sight vocabulary acquired by the three-year-olds and the kindergarten children was not impressive; but in the first case it was only late in the study that purposeful teaching of sight words was undertaken, and in the second case the amount of educational intervention was so minimal that little more could have been expected. Learning to recognize a certain number of words is quite a different matter from learning to read, however; children who accomplish only the former are known as "functional illiterates." It is reasonable to suppose that mastery of sight vocabulary will be of some value to a child in his later encounters with the real problem of decoding print, but this is only a conjecture.

The two studies that dealt with teaching logical operations both achieved considerable success, but it would be fatuous to claim that in the one case kindergarten children were moved up to the level of concrete operations and that in the other, preschool children vaulted to the formal level. It appears, rather, that what was learned was something specific and limited, even though in each case enough transfer was demonstrated to show that genuine conceptual learning took place rather than mere memorization of responses. Simple response learning is ruled out in the conservation study by the fact that the training involved judgments of difference whereas the conservation tests required judgments of sameness. Simple response learning is ruled out in the formal operations study by the fact that the test task involved a different combination of descriptive statements from those involved in the training tasks--only the formal structure of the statements was the same.

What, then was learned? Without going too far beyond the data, the most that can be inferred is something like the following: In the conservation experiment, the children learned that in making judgments of quantity appearances can be deceiving and so you should keep in mind what you already know about the amounts in question. In the formal operation experiment, the children learned a basic pattern for dealing with systems of statements having a certain logical structure. The kind of learning involved is much like that in learning a particular grammatical structure, like the passive voice construction, except that it involves a relationship among statements whereas the passive voice construction involves a relationship among sentence elements. If these inferences are correct, the learnings that took place in each experiment, while not trivial, were certainly quite limited in scope. In the context of a full curriculum for early education, these learnings would be but two minor items, ones which might not even be judged to be of sufficient importance, when compared with other possible items, to justify the time and effort required to teach them.

A Piagetian educator might well agree with this deflating judgment, saying that indeed nothing was taught but a couple of specific tricks, whereas what really counts is fostering overall progress up the ladder of cognitive development. But this would be begging the question, for it merely sets one level of abstraction against another. Real-life teaching does not occur in the abstract, but consists of actual behaviors and its effects can only be observed in actual behaviors. If what the children learn to do is of slight value, it is hypocritical to glamorize it with labels such as "cognitive growth" or "transition to the stage of concrete operations." A more profitable course would be to take the specific learnings for what they are worth and try to build a curriculum out of the ones that have some demonstrable importance, irrespective of their place in a classification of cognitive levels.

Thus, the pedagogical lesson to be drawn from the two studies of logical operations is, not that concrete or formal operational thought can be produced in preschool children, but that young children can be taught tasks that are classified at these levels. The two tasks used are merely representative of a multitude of such tasks, some of which might be harder to teach , some easier. Whatever its descriptive merits might be, a general classificatory scheme such as Piaget's cannot tell the educator whether or not children will be able to master any specific task. What is more, it provides no realistic guidance for judging the educational significance of any particular learning task. Hence it is difficult to see what value it can have for the planning of early education curricula.

The experiment on "creative" construction activities did not provide any empirical basis for judging the prospects for the promotion of learning in this area. Informal observation during the brief course of training sessions suggests that children can, at least, learn a number of useful procedures that would make creativity in construction a possibility and that they may be able to learn habits of self-questioning that will lead them to more complex and goal-directed activity. Although this is purely a matter of subjective judgment, the close observation of some 80 preschool children at work over periods of nine days each under a variety of conditions of freedom, restraint, and guidance, strongly reinforced the belief that the widely-acclaimed creativity of young children exists largely in the eye of the beholder; that left to themselves children tend to behave in a stereotyped and repetitive manner; that they seldom work toward any practical, conceptual, or esthetic goals in their productions, and when they do they are easily led astray by imitation, accident, or force of habit. Thus it would appear that the realistic task for early education is not to "preserve the spark" of creativity, but to provide the child with enough tinder in the form of skills and habits of thought that a spark will ignite if one should ever happen to occur.

To summarize, if we ask the question, "How much can the intellectual development of young children be accelerated?" the answer suggested by these studies is that there is no answer. If instead the question asked is "Can children be taught such-and-such specific thing?" the answer is very likely to be yes. What this tells us about children's potentialities is precisely nothing. It merely tells us something about the potentialities for figuring out instructional techniques to achieve definite goals. But what about raising the IQ or something global like that? This we did not try to do, and so we can only speculate. But if by raising the IQ one means teaching children to execute the tasks that are presented on the Stanford-Binet, for instance, one might generalize from the present results to predict that ways could be found to teach these tasks much as ways were found to teach word recognition, conservation of liquid quantity, and the solving of center problems, and that thereby the IQ could be substantially raised. Would this mean that dullards were being turned into gifted children? Not at all, no more than the four-year-olds who were taught to solve center problems were turned into the intellectual equivalents of twelve-year-olds. But to the extent that what the children were taught was useful one would have enabled them to behave more intelligently and to the extent that it was useless one would have wasted their time. From an educational standpoint, I do not see anything else that "raising intelligence" or "accelerating intellectual development" could mean.

Turning now to the matter of promising approaches to early education, three approaches were found to have some effectiveness: direct instruction, as in the teaching of sight vocabulary, construction skills, and formal operations; modeling, as in the teaching of conservation and in the guided problem-solving approach to construction activities; and the free use of teaching machines, as in the kindergarten reading experiment. Each of these approaches would seem to be capable of considerable development, and there is no evident reason why they should be incompatible with one another. Only direct instruction, however, has the scope and flexibility that make it clearly applicable to a large range of teaching tasks. Modeling and teaching by machine both require considerable ingenuity in designing appropriate ploys for each new task and there may be no way of adapting them to certain kinds of tasks. For conveying information or teaching clear-cut rules, modeling seems irrelevant. Teaching machines, on the other hand, have their most serious limitation in their inability to handle complex verbal responses on the part of the learner. Although computer technology is making progress in this direction, it is still a long way from even a crude approximation to the language-processing abilities of the human teacher, and crude approximations are not likely to be of much pedagogical value.

"Direct instruction," as the term is used here, includes the large range of normal teaching activities that involve demonstrating, telling, and asking. Thus we are talking about a time-honored approach that no doubt comprises the vast majority of teaching acts throughout the school years. It is, however, a novelty in preschool education. In our first study, the one on teaching reading to two- and three-year-old children, we drifted into direct instruction inadvertently, having assumed that it could not work with children so young. We were surprised at their willingness to accept it and their ability to profit from it. The major obstacle to its use, we conclude, has nothing to do with its effectiveness but rather with the fact that it requires some clear notion of what it is, specifically, that one is trying to teach. Early education has had little truck with specifics, except in the teaching of personal care and housekeeping practices. But if our earlier conclusion is accepted, that expectations for early education should be recast in the form of specific learning objectives, then it becomes reasonable to contemplate giving direct instruction a larger place in the preschool program. The next portion of this report is devoted to two experiments in the full-scale use of direct instruction at the preschool level.

3.0 Studies in Direct Verbal Instruction

The studies reported in section 2 represent something of a hodgepodge. We were interested in exploring different areas of children's learning to see what could be accomplished in each, and in trying different approaches to see what promise they held. In reporting these studies, after all of them had been completed, it was impossible to avoid bending the interpretation of them toward the conclusions that eventually emerged. Let me therefore confess that at the outset I, personally, was quite open-minded, which is to say naive, and had no inkling of where the studies would lead, As far as I was concerned, learning to read, acquiring the ability to cope with construction problems, and mastering logical operations were quite different matters, having nothing in common except that they were all components of intellectual growth. My views on teaching methods were all ad hoc. It did not occur to me that one might learn something from the teaching of logical operations that would be applicable to the teaching of reading, or vice-versa.

The expected follow-through on these exploratory studies was to have been more extensive and rigorous investigation of a few promising leads. The main conclusions that emerged from these studies militated against this strategy, however. We could, to be sure, have pushed ahead with more concerted work on the teaching of reading or constructive problem-solving or conservation. In no case did the results indicate a blind alley. Instead it began to appear that whatever we chose to teach could be taught to some effect, and that therefore the main problem was to determine what was most worthwhile to teach and to design a coherent program that would get the most educational mileage out of a given amount of instructional time. Thus, from the conducting of single-purpose experiments we were led to the designing of a multi-purpose educational program. Arrived at by this route, however, the program was quite different from other programs that originated from global aims. It was conceived of as an aggregate of hundreds of specific learning tasks, each of which had to be justified on grounds of instrumentality or intrinsic worth, rather than as derivatives of some general approach to the fostering of intellectual growth. Direct verbal instruction was chosen as the principal pedagogical method, not because of any presumed value as a form of experience for young children, but because, once the learning tasks had been sufficiently specified, there did not seem to be any reason to pursue them by more round-about means.

3.1 An academically-oriented preschool for disadvantaged children: results from the initial experimental group

3.11 Introduction

This report covers two years of work with a group of disadvantaged preschool children in which we tried to see how much the children could learn in the areas of language, arithmetic, and reading through straightforward teaching of carefully presented content. Our working assumption was that disadvantaged children differed from others only in what they had and had not learned previously, so that an effective compensatory education program could be designed simply by starting at a lower level of presumed initial knowledge and moving more rapidly and efficiently, thus enabling the children to catch up to others whose intial levels of knowledge were higher. The teaching program used in this study is evaluated on two bases: 1) its measurable effects on achievement and language test performance, and 2) its potentialities for further development to eliminate gaps and weaknesses revealed by the initial experimental group. Although we consider the immediate effects of the program to have been highly satisfactory, we consider that the second kind of evaluation is of the utmost importance in judging the long-range promise of various approaches to early education.

At the time of this writing, a second group of disadvantaged children has passed through the preschool program and is now in its second year, as is a group of middle-class children who began as four-year-olds in the same type of program. A mixed group of lower- and middle-class children is midway through its first year of preschool instruction. This report, however, deals only with the first group, which was treated with a different emphasis from subsequent groups. With the first group, our major interest was in getting some idea of what the limits of attainment might be for disadvantaged preschool children, and both the objective results and the post-mortem examination of potentialities for program improvement are approached with that interest in mind.

3.12 Method and procedures

Subjects: Subjects in the initial experimental group were recruited in the following way. In one elementary school in the Champaign-Urbana, Illinois, community, which served a predominantly Negro (over 90 per cent) lower-class population, all teachers made visits to homes of their pupils during the early fall. Within two weeks after these visits, teachers were asked to list students who (1) were having educational difficulties, (2) came from homes in which cultural deprivation appeared to be a major contributing factor, and (3) came from homes in which there was a four-year-old child. The lists were to be ordered, starting with the most serious cases. Members of the research staff first determined whether the four-year-old child was in the age range that would qualify him for entry into kindergarten the following year (four years old by December 1). They then informed qualified parents that a preschool was being established which would attempt to teach their child those things that would help him to succeed better when he started school. Of the first 18 cases so identified, there was only one parent refusal. The children were then brought to the University and administered the Auditory-Vocal Automatic and the Auditory-Vocal Association subtests of the Illinois Test of Psycholinguistic Abilities (ITPA). To be accepted, a child had to produce some intelligible verbal response (not necessarily correct) during the test. Two children, a pair of severely retarded and hyperactive twins, failed to meet this criterion and were therefore rejected. The remaining 15 children constituted the experimental group. There was no control group, the manner of selection being such that the children selected constituted a unique population subset.

Design: In the absence of a control group, the standard chosen against which to evaluate results was that of test norms. Having chosen a group which met rather stringent criteria of poor prognosis (without, on the other hand, having selected subjects on the basis of low test scores and thus inviting a strong regression effect), it appeared that bringing the children up to normal on appropriate tests would constitute a reasonable short-term criterion of success of the educational effort. The measures employed were the ITPA, the Stanford-Binet (Form L-M), and the Wide-Range Achievement Test. The following testing schedule was followed, timed by weeks from the beginning of instruction:

Minus one week	Auditory-Vocal Automatic (ITPA) Auditory-Vocal Association (ITPA)
Fourth-Fifth week	Full ITPA
Sixth-Seventh week	Stanford-Binet
Sixteenth week	Auditory-Vocal Automatic (ITPA) Auditory-Vocal Association (ITPA) Vocal Encoding (ITPA)
Twenty-ninth week	Full ITPA Stanford-Binet
Fortieth week	Wide-Range Achievement Test

The major pretesting was purposely delayed until the children had become well adjusted to school and had begun to make satisfactory progress in it. This is an important point to consider in judging the magnitude of gains from pretest to posttest, since the interval of schooling *preceding* the pretest in this study was as great as that *between* pretest and posttest in many studies of short-term intervention. These short-term studies, such as those that have since been carried out on Head Start programs, frequently show gains of six IQ points or more when pretests are administered before schooling begins. This point is not germane, however, to the more central issue of the absolute level of attainment that could be reached by the children.

Treatment Procedures

A. General

Classes were conducted for two hours a day, five days a week. With occasional exceptions for field trips, the following schedule was adhered to at all times, starting from the first day. The children were divided into three groups of approximately five each. There were three teachers, one for language, one for arithmetic, and one for reading. After an initial ten-minute period of free play, each group would go off to a different room for a twenty-minute class in one of the subjects. Then all would come together for toileting, snacks, and singing, all of which occupied thirty minutes. Following this each group would go off for another twenty-minute subject-matter class. They would then come together for a twenty-minute period usually devoted to reading and discussion of stories and then would separate for the final instructional session. The schedule was thus somewhat like a high-school schedule of separate subject-matter classes, but with a homeroom period intervening between each instructional session. All teachers participated in the "homeroom" activities. A fourth teacher worked remedially with children whose performance was too low to permit them to keep up with the instructional groups. Initially three children were in this group. By the end of the third month the number was reduced to one. The three regular groups were stratified according to level of performance, the initial grouping being made on the basis of ITPA scores, but with frequent shifts being made as performance levels changed. The only other modification was that for the first two weeks instructional sessions were limited to fifteen minutes instead of twenty, the "homeroom" period being expanded accordingly; twenty-minute instructional periods were instituted as soon as the children appeared able to maintain attention and effort for that length of time.

The instructional sessions were represented to the children as work rather than play, the child's responsibility being to speak when called upon to do so, to "try hard" to give the correct responses, and to refrain from diversionary activities such as social play or running around the room. Adherence to these behavioral rules was rewarded by verbal praise, fortified during the first month by cookies. Children were reprimanded for deviations from the rules and, if this was not effective, were excluded from the instructional group for short periods of time. Every effort was made to keep the instructional sessions lively and enjoyable and to shift the basis of motivation to the children's own accomplishments and progress as improvement became demonstrable.

The general instructional strategy followed in all three subjects was that of rule followed by application. A verbal formula was first learned by rote ("This ______ is ______," "If you start out with _____ and get _____ more, you end up with _____.") and then applied to a graduated series of analogous examples of increasing difficulty. Tasks were initially presented in a highly structured form that provided a maximum of syntactical and presentational prompts; then the task was systematically "destructured" to remove these prompts and admit the variations in presentation that would be encountered in normal situations.

B. Academic Curricula

1. Language. The language program was based on the following principles:

a. It attempts to focus on minimum essentials of language competence rather than allowing the content of the language program to be determined adventitiously. The minimum essentials were identified, not on the basis of frequency of use, but on the basis of the logical requirements of a communication system that will permit academic teaching to go on. The objective is a kind of basic English that teacher and child may use in the conduct of elementary education--a basic English, therefore, which does not embody all the concepts a child should master but which provides a medium through which those concepts may be learned.

b. Recognizing that learning the rules of language and logic is a matter of grasping and generalizing analogies, the program is structured so as to dramatize those analogies. Rather than grouping concepts on the basis of their thematic associations (concepts related to the school, to the zoo, etc.) they are grouped together on the basis of the rules governing their manipulation. Thus polar sets of diverse content (big-little, hot-cold, boy-girl) are taught as part of a single sequence, so that the child may eventually come to grasp the major principle governing such sets--the principle that saying that something is not one member of the set is equivalent to saying that it is the other member of the set.

c. Every effort is made to maximize the number of monitored responses that each child makes per class period. This, we feel, is the most critical tactical problem to be solved in teaching language to disadvantaged children--how to cram more "trials" into the limited time available for training.

To sketch briefly the broad outlines of the program, it presumes nothing more of the child at the outset than that he be capable of making some attempt at imitating what is said to him. It begins by teaching the basic identity statement, "This is a __________," and "This is not a ______________," applying it to familiar objects. Once this statement is mastered (and mastery of the not-statement is a major challenge to many seriously deprived children), the remainder of the beginning language program is devoted to work with the statement form, "This ________ is ____________," "This ______________ is not ____________," and the plural variations, introducing several different kinds of concepts that are used in the predicates of these statements: polar sets, as mentioned above; non-polar sets, such as the colors and prepositional phrases; and sub-class nouns, as in "This animal is a tiger." The length of time that it takes four-year-old children to master this very limited system varies considerably--from six or seven months for those who come in with practically no spoken language to two or three months for those of near-normal language competence. Once the basic system has been mastered, however, it has been found possible to move very rapidly with almost all children through the expansion of the system to include active verbs, the common tenses, and personal pronouns. The remainder of the language program is devoted largely to if-then type statements in which the major problems are logical ones concerning the use of all, only, some, and or. The program, thus, culminates in the use of language for deductive reasoning, all of the more elementary work with statement forms and concept types having been designed to provide the groundwork for this use of language.

In implementing the program, a few basic "moves" constitute most of the teacher's repertoire.

1. Verbatim repetition:

TEACHER: This block is red. Say it ...

CHILDREN: This block is red.

2. Yes-no questions:

TEACHER: Is this block red?

CHILDREN: No, this block is not red.

3. Location tasks:

TEACHER: Show me a block that is red.

CHILDREN: This block is red.

4. Statement production:

TEACHER: Tell me about this piece of chalk.

CHILDREN: This piece of chalk is red.

TEACHER: Tell me about what this piece of chalk is *not*.

CHILDREN: (ad lib) This piece of chalk is not green...not blue, etc.

5. Deduction problems:

TEACHER: (with piece of chalk hidden in hand) This piece of chalk is not red. Do you know what color it is?

CHILDREN No. Maybe it blue...Maybe it yellow...

These moves represent a rough hierarchy of task difficulty. In early stages of the program, large amounts of time have to be devoted to the lowest level--verbatim repetition--and deduction problems can seldom be handled. By the end of the program, most of the time is devoted to deductive problems, although at each new step in the program it is necessary to go through all of the moves, if only in very condensed form.

2. Arithmetic. Since the arithmetic of natural numbers is reducible to counting operations, it can be taught legitimately as a kind of "science of counting," without reference to the multitude of real-life phenomena that can be interpreted arithmetically. This was the approach taken in the arithmetic program. It is justified on the grounds that the disadvantaged children we were working with lacked the verbal and logical sophistication necessary to abstract arithmetic principles from the phenomenal "noisiness" of everyday activities in the manner required by "activity methods" of teaching arithmetic. The introduction of this phenomenal noise was treated, instead, as a part of the gradual destructuring of tasks, described previously.

After the initial teaching of counting, arithmetic itself was taught through the use of equations, the key idea being that any equation could be read as a statement of fact and also as an instruction that told how the fact could be established through a counting operation. Thus, the equation, $3 \times 4 = 12$, could be read as the statement of fact, "Three times four equals twelve," and as the operational rule, "If you count by three's four times, you end up with twelve." The introduction of an unknown, as in $3 \times b = 12$, created a question, "Three times how many equals twelve?" and the operation for finding the answer created a question, "Count by three's how many times to end up with twelve?" Analogous statements and operations were used for addition, subtraction, and division (expressed by fractions).

The kind of pattern drill that was used in the language program to teach basic grammatical rules was used in arithmetic. Just as , in language, it is too much to expect the child to understand an explicit rule such as "When the subject is plural, change _this_ to _these_ and add _s_ to the noun," it is too much in arithmetic to expect the child to understand the explicit rule, "Adding zero to a number leaves it unchanged." Instead the child learns to operate in accordance with the rule by drill on a sufficient number of analogous instances that the rule generalizes to new instances "This box..., These boxes..." "This book..., These books..." etc. and similarly, "One plus zero equals one," "Two plus zero equals two," "Eight plus zero equals eight," etc. Pattern drills were used to teach the "plus zero" rule, the "plus one" rule, and finally the operation for working up from a given number-plus-zero to the given number-plus-a-given-addend: "What's eight plus three? If eight plus zero equals eight, eight plus one equals nine, eight plus two equals ten, and eight plus three equals eleven." Ample use of concrete demonstrations was made, of course, to show the validity of these rules.

3. _Reading_. The program that was offered in reading must be understood in the light of our expectations at the time as to how far we should be likely to get in teaching reading to young disadvantaged children. Our previous work with more advantaged, though younger, children and the poor record of achievement of disadvantaged children in reading had not left us very optimistic on this account; at the same time we recognized that the usual kinds of training in reading skills had little relation to and would therefore probably have little transfer to other kinds of learning. We are referring here to the mechanics of reading, to the rules for decoding printed characters into spoken words--not to the aspect of reading which involves comprehension

and for which the language program seemed to provide adequate preparation. We therefore decided to attack the mechanics of reading in a way that might not be the most productive for the immediate purpose but which would give the children experience in working with an explicit rule system similar to the systems they would encounter in other formal disciplines.

By introducing certain artificial restrictions, we were able to reduce the inconsistency and complexity of English orthography and highlight its logical aspects. We restricted the initial vocabulary to three-letter consonant-vowel-consonant patterns, and avoided use of some of the more troublesome consonants. For further simplification we used only lower-case letters. Within this restricted set of words, the following six sets of rules hold:

1. A word has a beginning and an end. (The beginning is the initial consonant; the end consists of the vowel and the final consonant.) If it has a beginning and an end, it is a word. The beginning always comes before the end.

2. If the beginnings are the same and the ends are the same, the words are the same and they look the same.

3. If the words are the same, they sound the same. If the words sound the same, they are the same.

4. If the words look the same and sound the same, they are the same.

5. If the endings of the words are the same, the words rhyme. If the words rhyme, the endings are the same.

6. If the beginnings of the words are the same, the words alliterate, if the words alliterate, the beginnings of the words are the same

Parts of words were presented to the children on cards, the initial consonant (the beginning) on a white card and the last two letters (the ending) on a yellow card.

Learning to apply the rules required, of course, learning the implied visual discriminations ("look the same") and auditory discriminations ("sound the same"). Learning this set of rules and learning the conventional sound values of the alphabet was taken to constitute the readiness phase of reading instruction, after which the program proceeded with a rather conventional phonic approach, using spelling patterns that followed the order of Bloomfield and Barnhardt's Let's Read (1961).

C. Other Activities

The three academic courses described in the preceding section took up half of the children's two-hour school day. The other half was less structured and intended to amplify and reinforce what was learned in the academic courses rather than to achieve additional objectives. The first ten-minute period was almost completely free and was generally used by the children for working puzzles, playing with a miniature house, furnishings, and family, or in casual conversations with teachers. Snack time was brief and the only teaching involved was concerned with identifying the color of the fruit drink served, which varied from day to day. In the singing period, which lasted fifteen to twenty minutes, however, songs were specially written and scheduled to provide practice in language operations that were being dealt with at the time in the language course--use of singulars and plurals, classification ("If it's a truck, then it's a vehicle," sung to the tune of "Old Gray Mare"), and reversal of elements in phrases. Because of the importance attached to this practice, children were required rather than merely encouraged to sing. Story-telling was also geared to the language program and involved a great deal more question-and-answer activity than is common in reading stories to children.

D. Atmosphere

Because the school was definitely teacher-directed and task-oriented, the above descriptions may evoke the image of the grim "school marm" dominating a group of surly and joyless children. However, this was seldom the case. The teachers thoroughly enjoyed teaching, learning, and working with children, and the prevailing atmosphere among both children and teachers was one of high spirits that tended to bedazzle observers who were accustomed to the more somnolent atmosphere of most nursery schools (see Pines, 1967).

3.13 Results

This section will be confined to analysis of objective test results. The necessarily more anecdotal and impressionistic observations of learning will be left to the discussion section. The group remained intact during the second year and received a program that was in most respects a continuation of that described in the preceding section, with differences that will be noted where they are germane to the results.

ITPA and Stanford-Binet

The test data will be presented graphically in a manner developed from suggestions of Hively (undated). Since the design of the study was not one that lends itself to the fruitful application of inferential statistics (no control group, exhaustive sampling of a particular population segment), this method was chosen to provide a maximum of descriptive detail and at the same time to depict the salient features of the score distributions. Age-level scores are used throughout, but these are adjusted by adding or subtracting a constant value to compensate for differences in chronological age. Thus, at the time of the initial pretest, the average age of the children was 4-6 (four years, six months). Suppose that at that time a particular child's chronological age was 4-8 and that on the Auditory-Vocal Automatic subtest of the ITPA he obtained an age-level score of 3-10. A constant of two months would be subtracted from his score, so that on the graph he would be shown as having an age-level score of 3-8, placing him 10 months below average, as was in fact the case. The same constant would be subtracted from all his other scores on other occasions and other tests. In this way, every child's deviation above or below his chronological age on a given testing can be read as a deviation from the same line. Referring to Figure 3.13.1, child D is seen to have scored 15 months below her chronological age on the first administration of the Auditory-Vocal Automatic subtest, 11 months below her chronological age on the second administration (which occurred about a month later, as shown by the second heavy vertical line), 2 months below on the third testing, 29 months above on the fourth testing, and 17 months below age level on the fifth testing. In this particular case, the age level score on the fifth testing was the same as on the third, but this of course represents a loss in relation to chronological age, because the two testings occurred 15 months apart. The subjects are arranged in descending order according to their scores on the fourth testing, which occurred at the end of the preschool year, the point of major interest in this study. A descending line connects scores obtained on the fourth testing for all subjects, making it possible to see at a glance how many children scored above or below their age level at that time (in this case, six above and nine below); the spread of scores (from almost two and a half years above average to almost two and a half years below average); and the central tendency of scores, represented by the balance point of the line, which is a couple of months below average. Similar lines connect scores at the first, second, and fifth testings, but they are less regular in configuration (because the fourth testing was used as the basis of ordering). Scores obtained on the third testing are not connected, since, with the amount of fluctuation in scoring that occurred, this would make the graphs too difficult to read.

For the reader who is unfamiliar with this method of presenting data, it may take a little while to become accustomed to reading the graphs; but the writer agrees with Hively that for descriptive purposes it represents a considerable advance over the usual summary statistics in that it enables the reader to consider the data according to different criteria of success and failure without the need for reanalysis. There is no reduction of the data. Individual scores can all be recovered. The constants used for each child to correct for his chronological age are as follows:

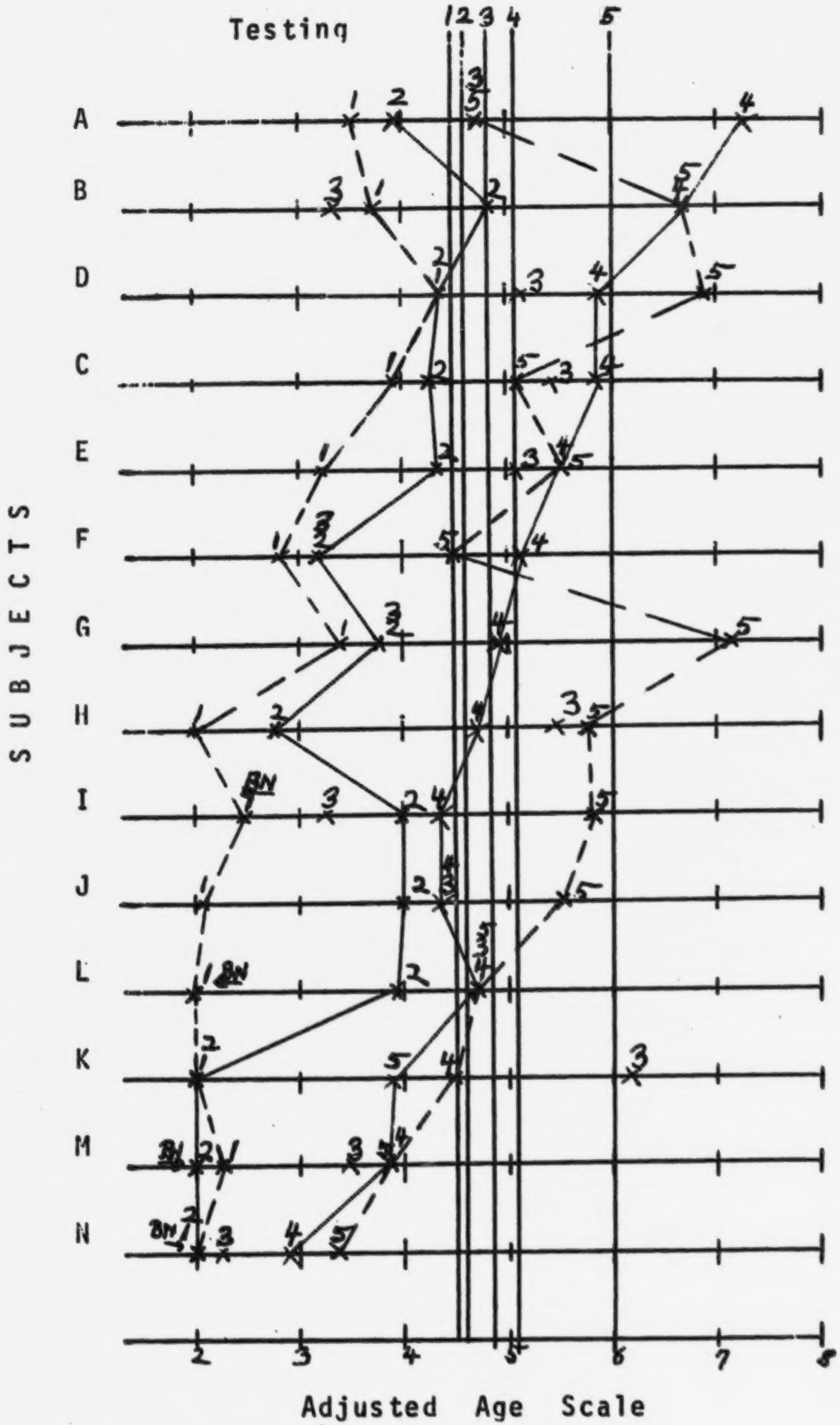
A + 1
B - 4
C + 7
D + 5
E - 1
F + 6
G + 1
H - 3
I - 3
J - 4
K - 4
L + 5
M - 1
N - 4
O 0

Auditory-Vocal Automatic (Figure 3.13.1)

This is a test of the child's ability to use grammatical inflections to indicate number, tense, and degree. It is perhaps the only subtest of the ITPA that can be interpreted, on the basis of content validity, as an achievement test; and although the language program was not primarily devoted to this aspect of grammar, many of the exercises did in fact deal directly with the production and comprehension of grammatical inflections. Not one child was scoring up to age level on this test before school began, the center of the distribution being about 18 months below average. By the second testing, a month after school began, nine of the children had come up to within a year of average, but only one scored above his age level. By the end of the preschool year, however (test 4), six of the children were scoring above age level, while only three children were more than a year below average.

During the second year, formal instruction in language usage was discontinued, being replaced by a course focussing on science and verbal reasoning. As the scores for the fifth testing show, only a few children made appreciable progress in mastery of grammatical inflections during the second year, half of the children registering no gains or regression. In retrospect, the result is not surprising, but it shows a clear inadequacy in the second year's program. To advance much beyond the five-year level on this test, a child must begin to master irregular plurals and tense forms. Since these are typically not used in his home environment, there is little way that he can learn them except through schooling; and as these results indicate, he will apparently not learn them unless they are deliberately taught.

Figure 3.13.1--ITPA Auditory-Vocal Automatic Scores on 5 Testings



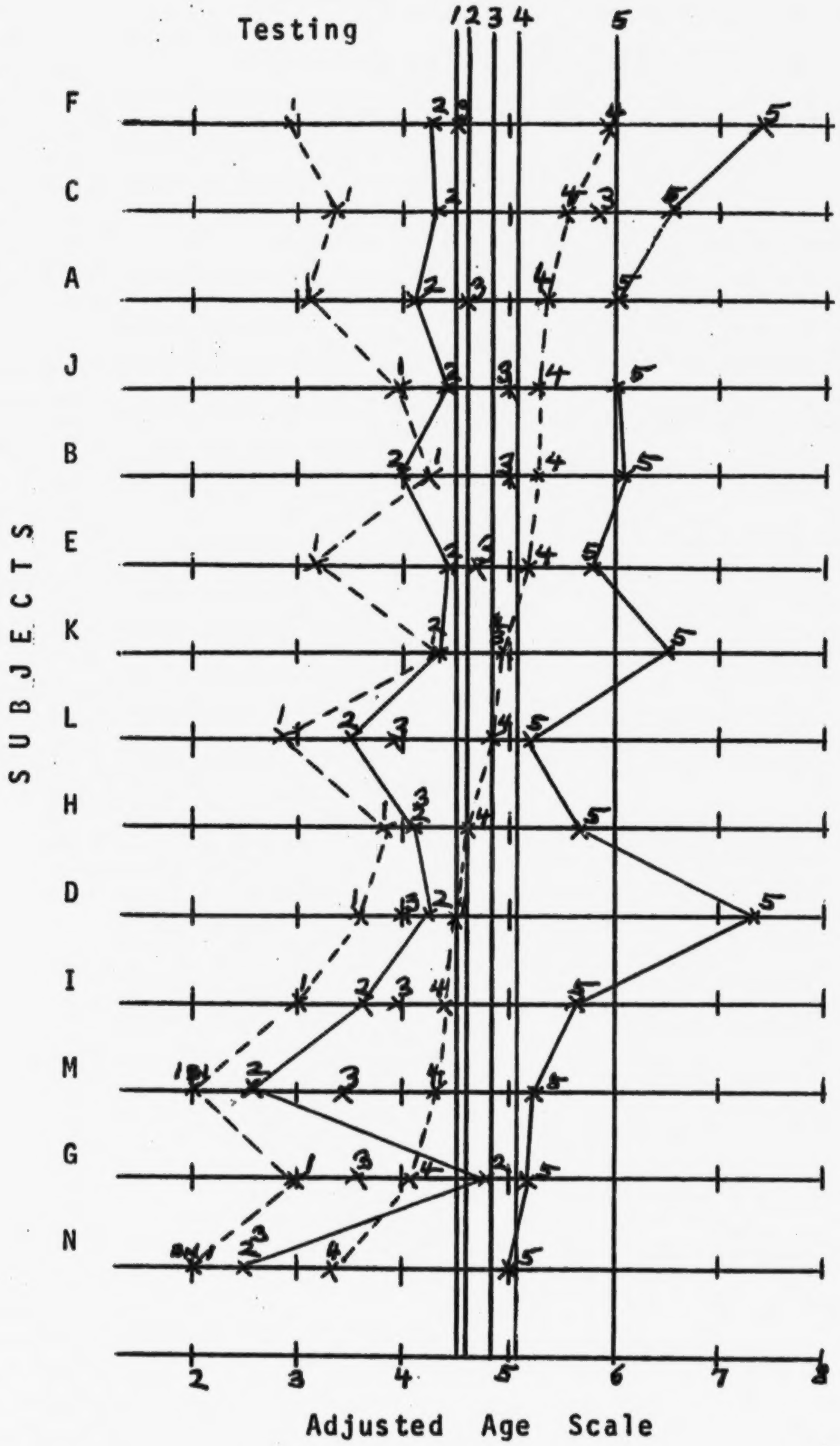
Auditory-Vocal Association (Figure 3.13.2)

This is a verbal analogies test ("During the day we're awake; at night we _______"). As a test of verbal reasoning it is not quite what one might wish for. Most of the items, like the one given, can be answered correctly without grasping the analogy, merely by completing the second clause in the most obvious way. Even so, it provides a useful indication of the child's ability to comprehend declarative sentences. Few of the sentences were of the forms used in the language program, and so the test may be taken as one of transfer to the diverse range of simple declarative sentences. On the first testing, as on the second, only one child scored up to his age level on the test. However, on the first testing 10 of the 15 children were over a year below their chronological age, whereas a month later this was true of only four children. By the fourth testing, at the end of the preschool year, five children scored above their age level, and only three were a year or more retarded. On this test, progress continued the second year. Six children scored above age level by the end of the second year, and only one of the children was over a year retarded. The group as a whole was approximately up to average. This makes sense in light of the extended emphasis on logical use of language in science and problem solving.

Vocal Encoding (Figure 3.13.3)

This test was of interest because it involves a "divergent" as opposed to a "convergent" use of language. Instead of there being one correct response for each item, each item called for the child to "tell me all about" an object that was presented to him, and the child's

Figure 3.13.2--ITPA Auditory-Vocal Association Scores on 5 Testings

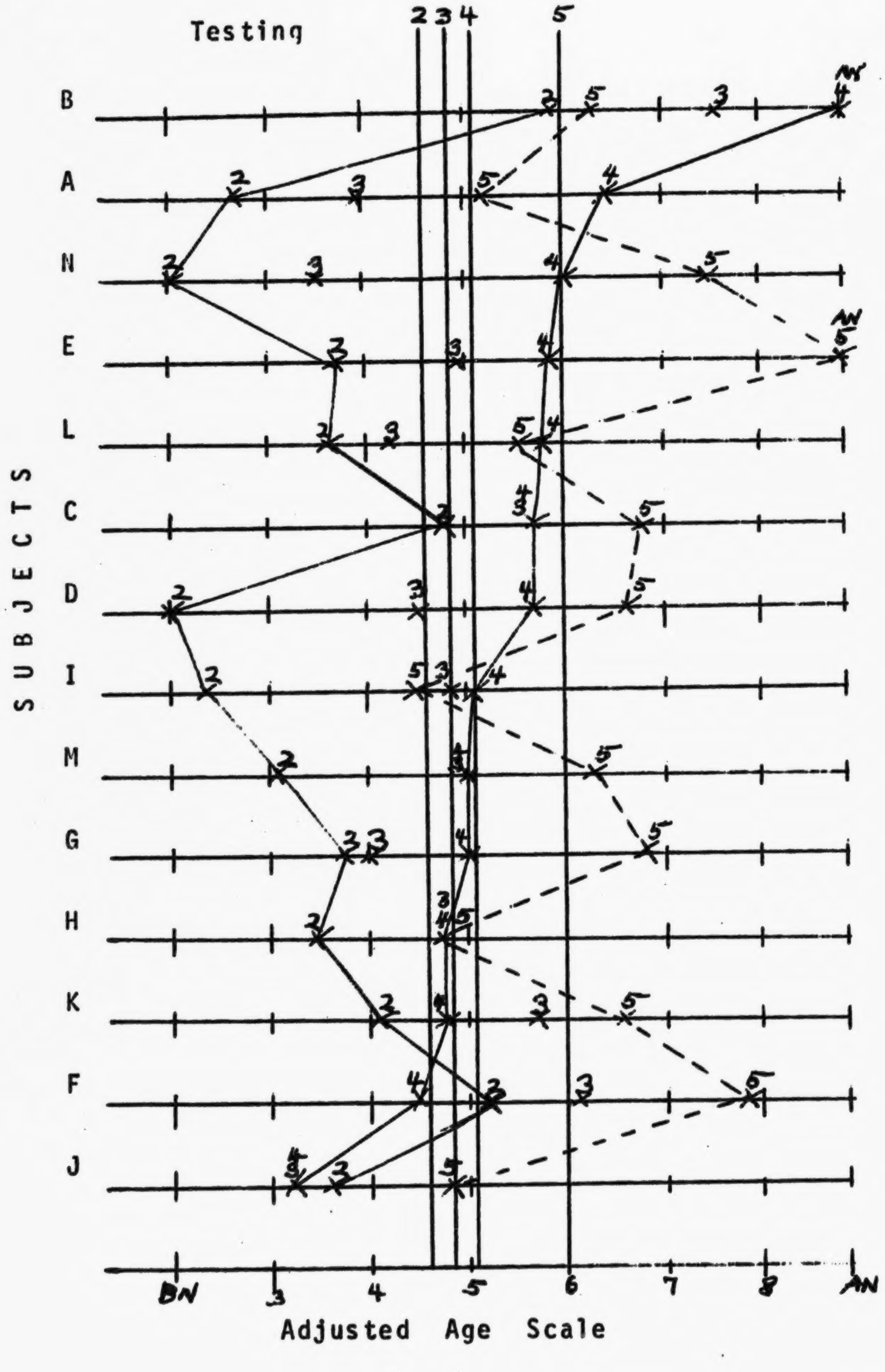


score was the total number of different appropriate things he said about the objects. This subtest, as well as the remaining ITPA subtests to be discussed, was not administered prior to any schooling, but was first administered after a month of school. At that time only three children scored at or above their age level as shown in Figure 3.13.3. Three months later (Testing number 3), the children had come up to age level in performance, seven of the children scoring at or above their age level. By the end of the year this number had increased to eight, and by the end of the next year it had increased to nine. Individual performances fluctuated a great deal from testing to testing, however. Performance on this test would seem to depend to a considerable degree on the motivational state of the child at the time of testing.

Auditory Decoding (Figure 3.13.4)

This is a vocabulary test in which the children respond yes or no to questions of the form, "Do lanterns shine?" Although vocabulary building was not a major objective of the program, it was nevertheless puzzling to discover that the children made practically no gain on this test during the first year. As Figure 3.13.4 indicates, however, this was largely a function of the surprisingly high performance of the children on their first testing--surprising in that large vocabulary deficits are expected in children of this kind. A subsidiary analysis of the test content and scores suggested that their performance may have been due to a floor effect on the test. The first twelve items of the test involve very simple words, so that errors would more likely be due to inability to understand the question form than to ignorance of the words. After item 16 the words increase sharply in difficulty.

Figure 3.13.3--ITPA Vocal Encoding Scores on 4 Testings



Getting the first 12 items right yields an age-level score of 4-1, close to normal for the first testing. All children got almost all of the first 12 items right, so that with a few additional chance successes they could earn normal or near-normal scores even with very weak vocabularies. The small gains obtained during the first year were due to getting one or two more items correct beyond the sixteenth item, rather than by getting a larger percentage of the earlier items right. The larger gains obtained during the second year lend further weight to the premise that during the first year the children were still at the artificial floor of this test. Had the children not already acquired sufficient language skill by the time of the first testing to handle the easy-word items, larger gains could have been expected during the first year; but it would have been a mistake to attribute these gains to vocabulary growth.

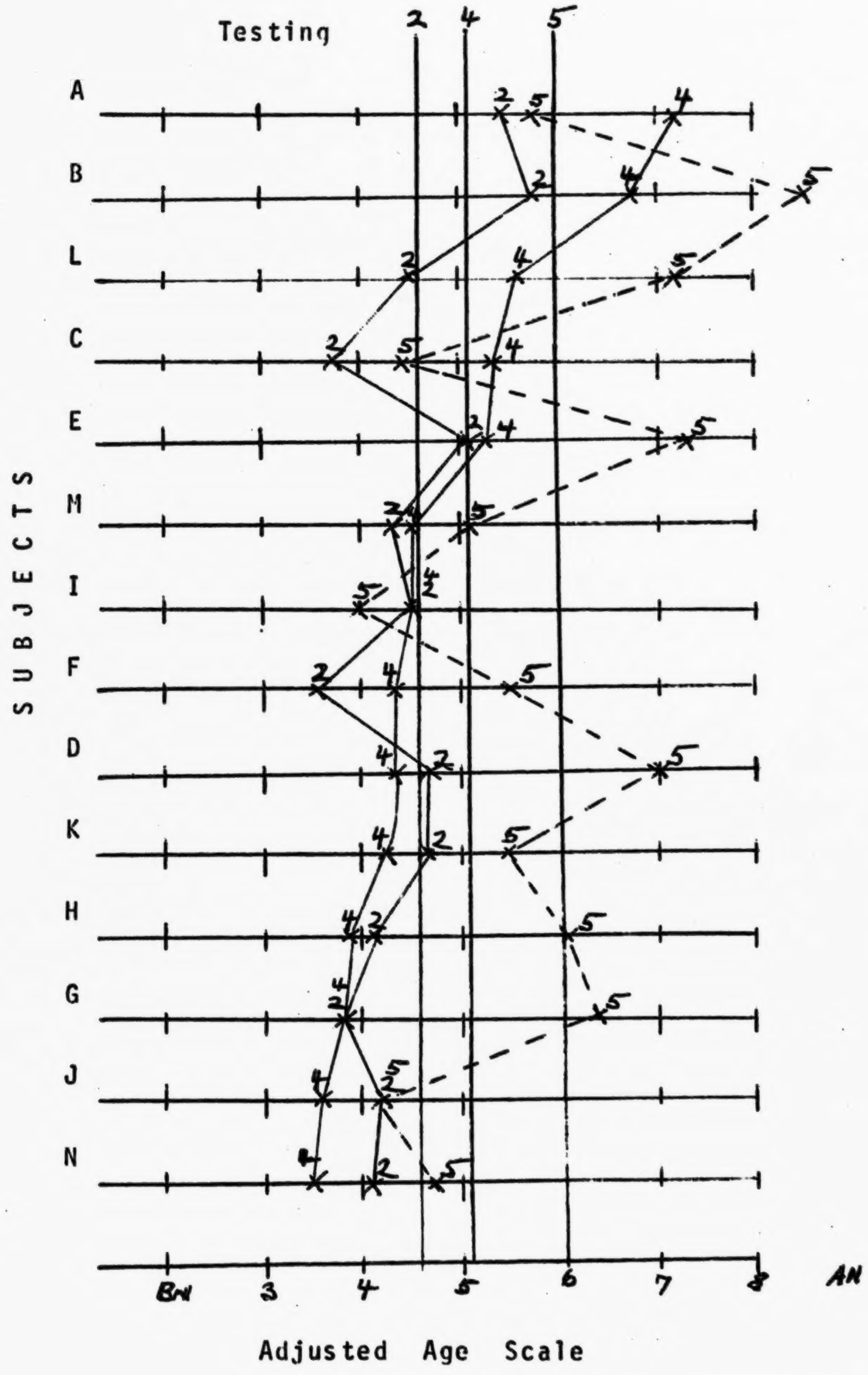
Other ITPA Subtests (Figures 3.13.5 - 3.13.9)

The ITPA subtests discussed above were of special interest because their content could be rather easily characterized and had some obvious relevance to school tasks. The remaining subtests are much less clearly interpretable on the basis of their content, and in the absence of validity studies,[1] their relevance to other kinds of behavior is not known. Two of the subtests, Auditory-Vocal Sequential (Figure 3.13.5) and Visual-Motor Sequential (Figure 3.13.6) are short-term

---

[1] In a validation study nearing completion at this writing, by Alfred Hirshoren, now in the Department of Special Education, DePaul University, four ITPA subtests were found to contribute significantly to prediction of achievement at the beginning of second grade from data obtained in kindergarten. They are Auditory-Vocal Automatic, Auditory-Vocal Association, Auditory-Vocal Sequential, and Visual-Motor Sequential.

Figure 3.13.4--ITPA Auditor-Decoding Scores
on 3 Testings



memory tests, the first calling for repetition of digit series and the second for reproduction of series of figures. It will be observed that performance on the auditory test was good to begin with and got progressively better, while performance on the visual task was poor and got relatively worse. To the extent that preschool training may have had anything to do with these trends, the difference would not appear to be due to the lack of work with sequences of visual stimuli, since there was a great deal of this in both reading and arithmetic, but in the much greater emphasis on reproducing sequences orally rather than motorically through writing or arrangement of objects. Visual Decoding and Visual-Motor Association involve rather similar tasks. In the first the subject is presented with a picture of an object and then must find, in a set of four pictures, a variety of the same object. In the second, the task is the same except that the subject must locate an object that "goes with" the sample object (such as nail, with hammer). On both of these tasks, as shown in Figures 3.13.7 and 3.13.8, the children rose to an above-average level of performance by the end of the first year, but performed at about an average level by the end of the second year. The final test, Motor Encoding (Figure 3.13.9), requires the subject to demonstrate through actions what one does with various pictured objects. Performance was poor during the first year and improved markedly for some children but not at all for others during the second year. The task would appear to have two unrelated components--one a familiarity with the objects depicted (household implements, musical instruments, etc.) and the other an aptitude for pantomime. A case-by-case analysis of the widely varying scores on testing number five suggested that the children had, by the end of the second year, acquired a sufficient amount of familiarity with the kinds of objects depicted, but that the more physically inhibited children failed to demonstrate this on the test.

Mental Age (Figure 3.13.10)

The most global and therefore least interpretable measure of change was that afforded by scores on the Stanford-Binet Intelligence test. These scores, being more reliable than those on ITPA subtests, showed greater stability from testing to testing than the others. A fairly steady and consistent shift of scores from slightly below average to slightly above average is indicated. What is perhaps most noteworthy is that the gains were shared about equally by all children. There was only a slight suggestion of regression effects at the top and bottom of the score distribution.

Achievement

The Wide-Range Achievement Test was administered once at the end of the preschool period (August, 1965) and again at the end of the kindergarten year (May, 1966). On the second administration, the test was administered in the prescribed way throughout. On the first administration, however, certain modifications were introduced to adjust to curriculum conventions. Specifically, the first eight items of the reading test, which consisted of naming the letters of the alphabet, were changed to present lower-case instead of upper-case letters, and credit was given for producing the phonic equivalents that the children had been taught instead of the conventional names. On the second administration the capital letters were given and only the appropriate letter names were counted correct. On the first administration of the arithmetic test problems were presented in equation

Figure 3.13.5--ITPA Auditory-Vocal Sequential Scores on 3 Testings

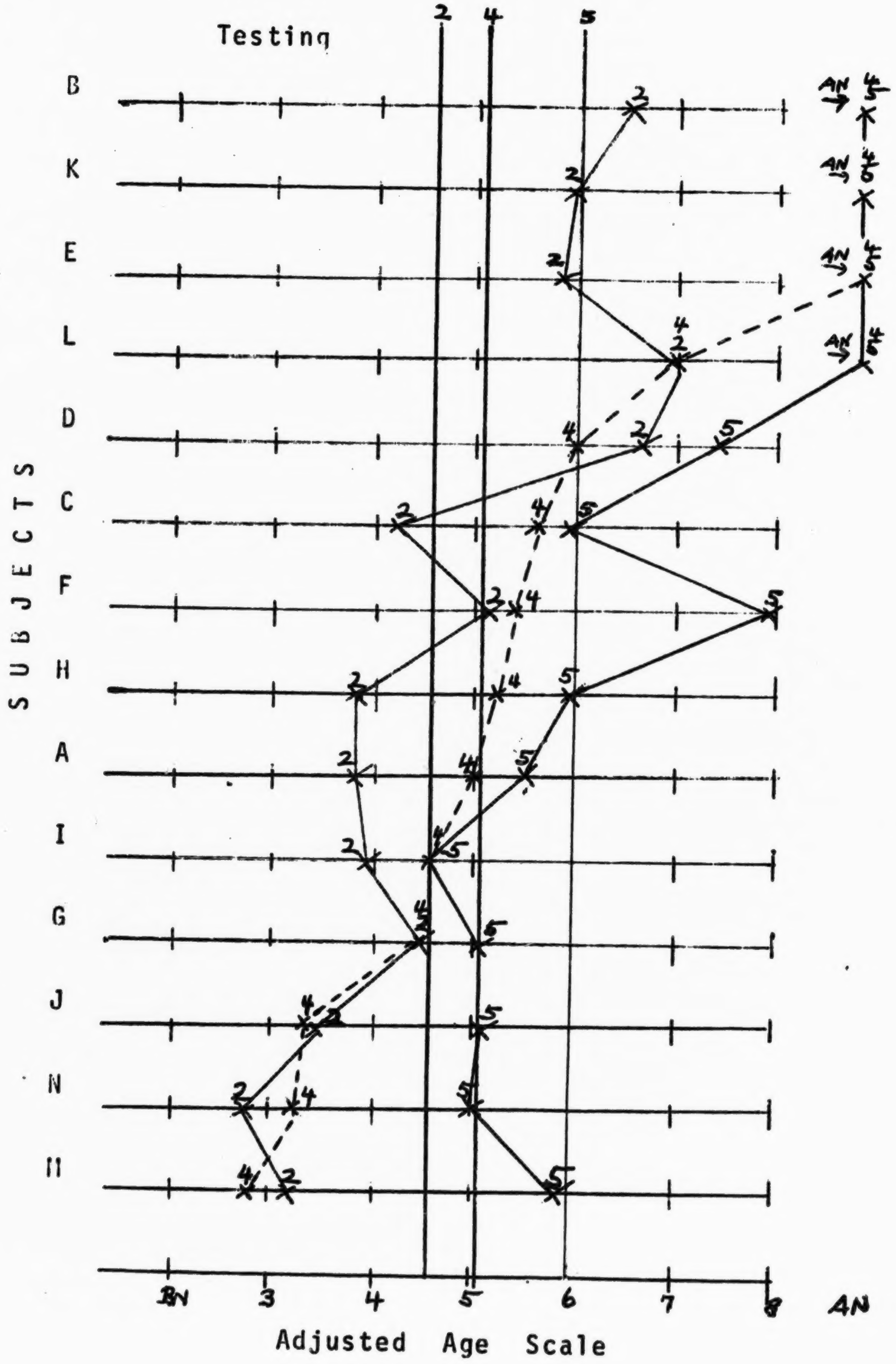


Figure 3.13.6--ITPA Visual-Motor Sequential Scores on 3 Testings

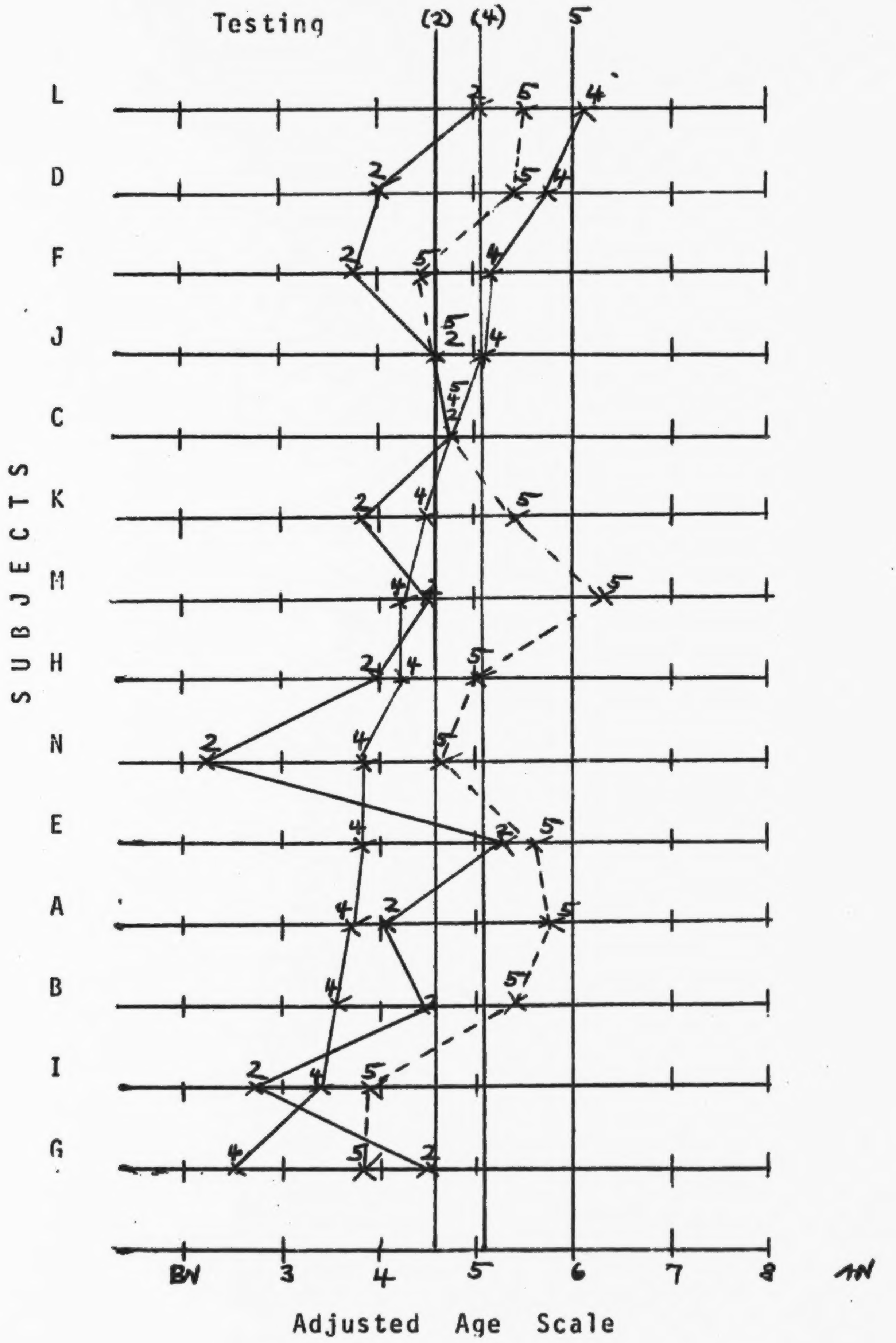


Figure 3.13.7--ITPA Visual Decoding Scores on 3 Testings

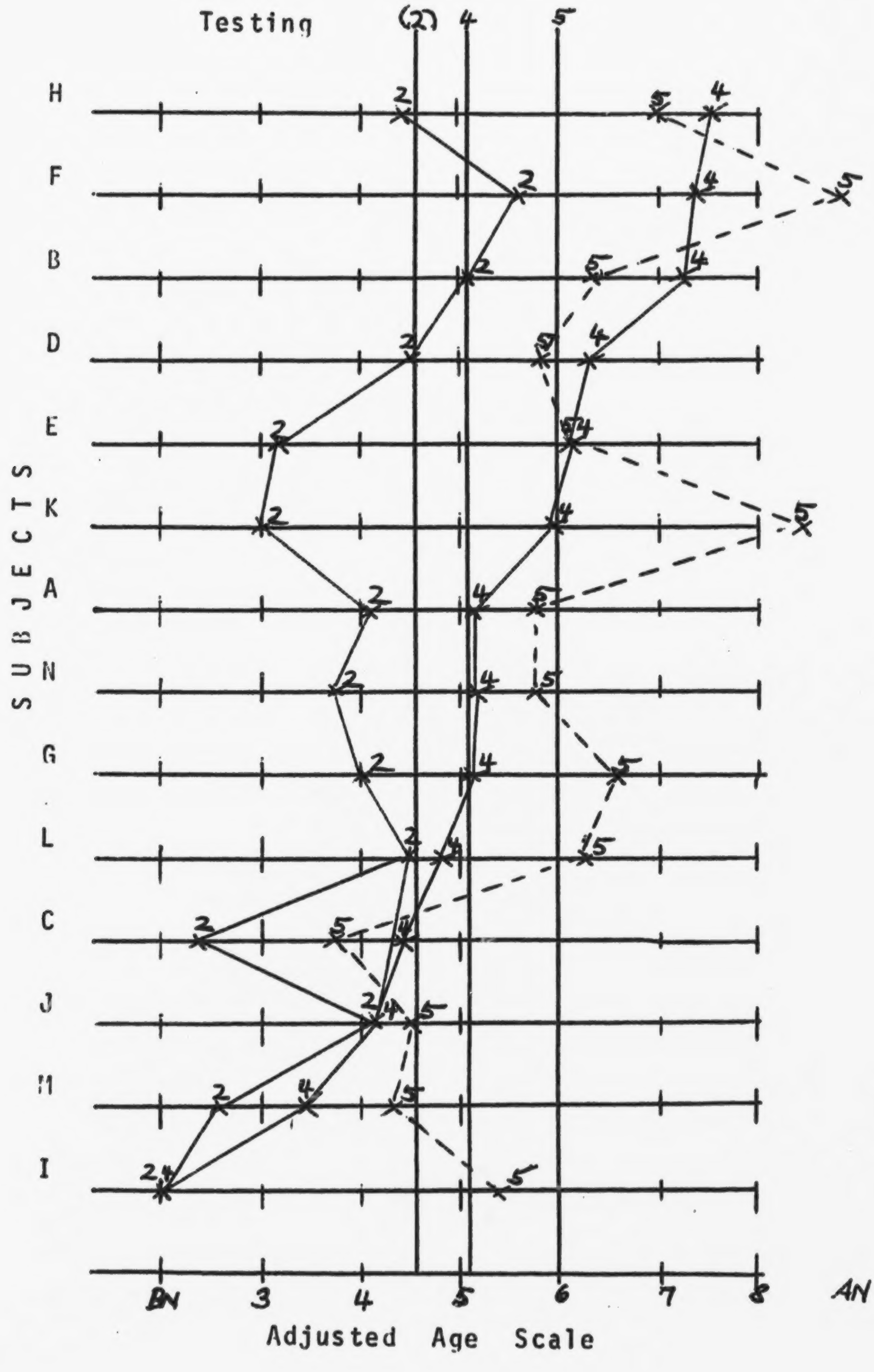


Figure 3.13.8--ITPA Visual-Motor Association Scores on 3 Testings

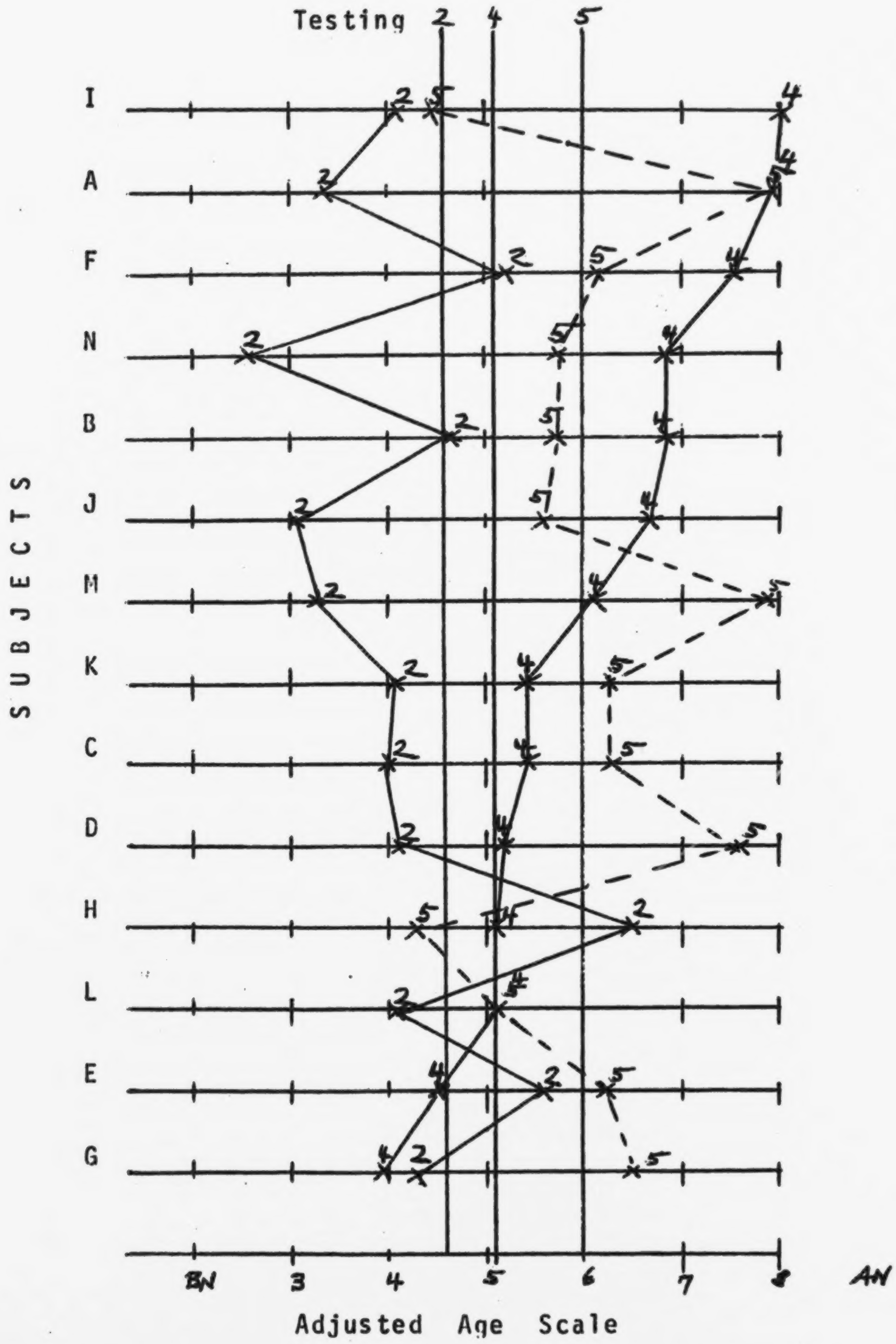


Figure 3.13.9--ITPA Motor Encoding Scores on 3 Testings

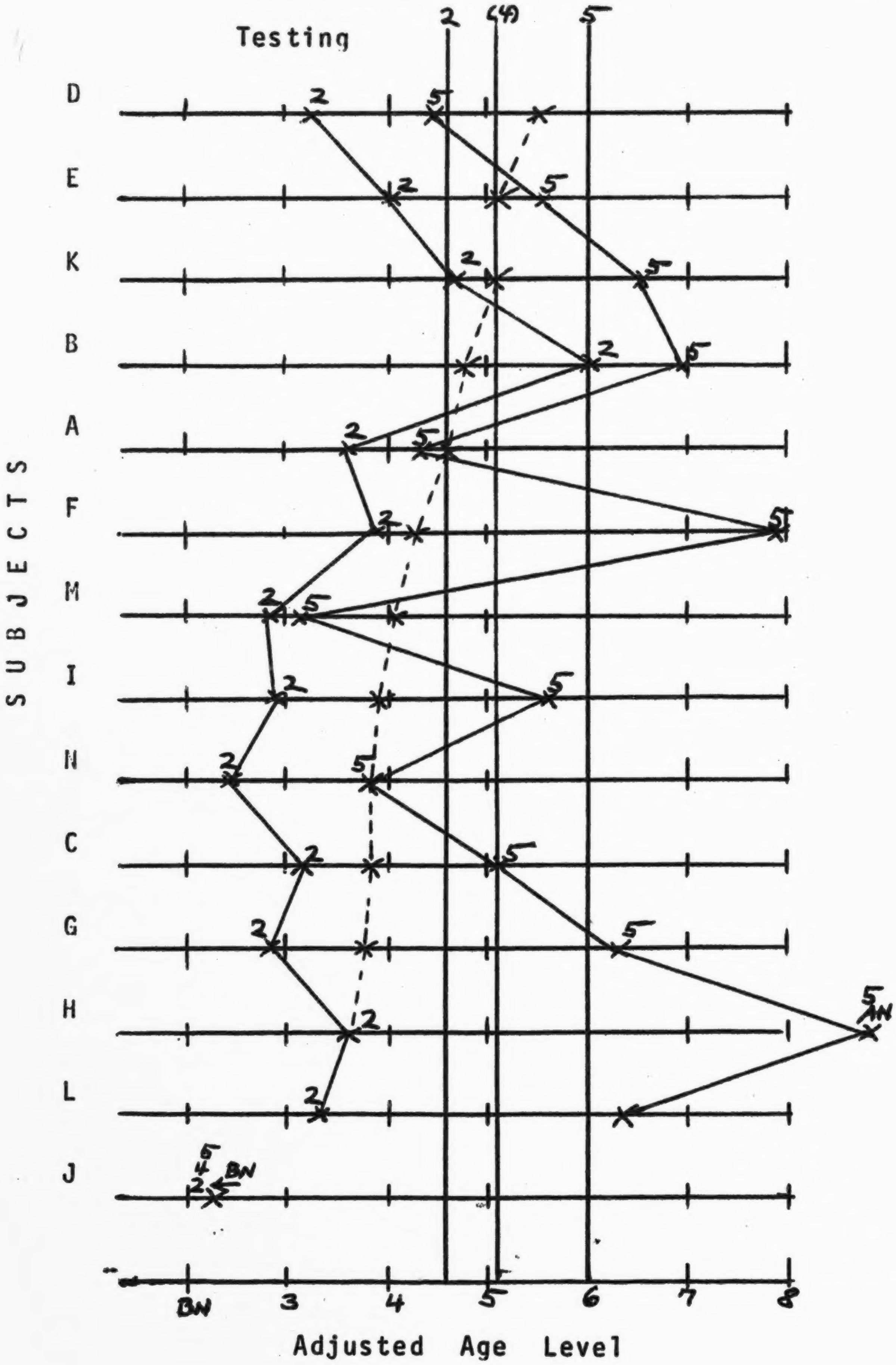
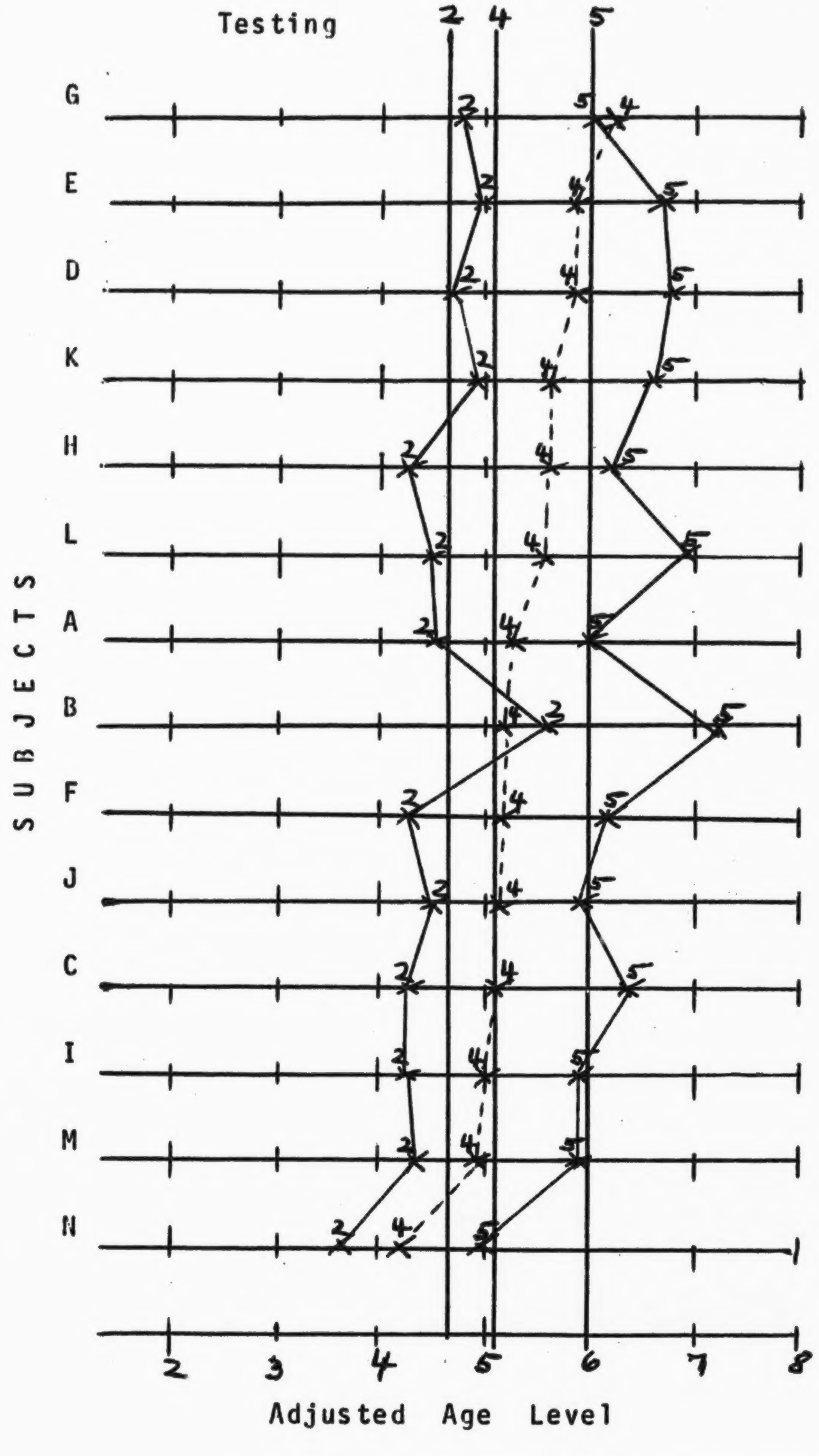


Figure 3.13.10--Stanford-Binet Mental Age Scores on 3 Testings



form (4 + 2 = b, etc.) and the children were allowed use of a multiplication table. On the second administration the problems were presented in vertical form, without operation signs (a form with which the children had had only limited experience, especially those in the lower groups) and the children were allowed no computational aids. As a result, scores from the first administration are not precisely comparable to the norms or to the scores from the second administration, whereas scores from the second administration should be fully comparable to norms. The reason for deviating from prescribed procedures on first testing was that we were interested in getting an interpretable picture of the children's learning, with some norm reference, and felt that if we followed standard procedures at that time scores would be seriously distorted by artifacts of test format. By the time of the second administration the children had been sufficiently exposed to the conventions of the test format that we believed reasonably valid scores could be obtained with it.

Because of the discrepancies in administrative procedures, we are not presenting the scores in graphic form as with the other tests. The scores are reported in full, however, in Table 3.13.1. By the end of the preschool year the average achievement of the children (not allowing for non-standard testing procedures) was at the beginning first-grade level in reading and at the beginning second-grade level in arithmetic. Spelling ability was low, however, only four of the children scoring at the beginning first-grade level. Printing and spelling were not taught during the first year; the four high-scoring children were ones who had received help at home, at least to the extent of being taught to print their own names. Systematic training in spelling was carried on throughout the kindergarten year, however, amd was reflected in such an increase in scores that there was no overlap between pre-kindergarten and post-kindergarten scores in spelling.

Table 3.13.1 - Individual and mean grade-level scores on Wide-Range Achievement Test

	Reading		Arithmetic		Spelling	
Child	Aug. 65	May 66	Aug. 65	May 66	Aug. 65	May 66
A	.5	1.3	2.3	1.8	.8	1.7
B	1.0	1.3	2.0	2.5	.1	2.0
C	1.4	1.7	2.7	3.6	1.0	1.9
D	1.2	2.1	2.2	3.6	1.0	2.2
E	1.1	1.9	2.5	3.3	.1	1.8
G	1.1	1.8	2.5	3.1	1.0	2.0
H	1.1	1.3	2.3	2.3	.3	1.7
I	1.0	1.3	2.2	2.5	.1	1.2
J	1.3	2.2	2.7	3.1	1.0	2.2
K	.9	1.3	2.0	2.9	.7	1.9
L	1.0	1.0	2.0	1.4	.5	1.2
M	1.0	.9	1.8	1.8	.0	1.3
N	1.1	1.1	.3	2.0	.1	1.5
Mean	1.05	1.50	2.12	2.61	.52	1.74

By the end of the kindergarten year mean achievement was at the mid-first-grade level in reading and the mid-second-grade level in arithmetic. The gain over pre-kindergarten scores was half a year in each case, but these gains cannot be taken at face value because of differences in testing procedure. In reading, for instance, three children, showed no gain (a slight loss in one instance). These three children, however, comprised the lowest reading group, in which capital letters were hardly treated at all. Inspection of test protocols showed that on the first administration (in which lower-case letters were used), these three children earned their entire scores of five or six points by letter reading. On the second administration (in which capital letters were used) each of the children got only two letters correct, the difference being made up by their reading some words correctly. If, as is permitted in the instructions for administering the Wide-Range Achievement Test, credit is given for all the letters, providing the child reads some words, then all but one child showed a gain (and the mean grade level is raised .1 years.)

3.131 Summary of test results

For those tests that had some clearly identifiable content, scores tended to reflect rather faithfully the amount of emphasis that was given to that content in the program. When grammatical usage was deliberately taught, scores rose dramatically on the test of grammatical usage (Auditory-Vocal Automatic); when instruction was terminated, improvement in scores ceased. Before the commencement of direct instruction in printing and spelling, the children's achievement in this area was negligible, even though they were learning some reading at the time. When direct instruction in spelling was instituted, scores rose sharply to the point where they were normatively as good as scores in reading. If there was any general improvement in test scores that could be attributed to "stimulation" or "enrichment," it was very slight compared to the specific effects and was largely confined to the first year.

A major interest of the study was to see how far disadvantaged children could be taken in selected areas of instruction through intensive and structured efforts. The results of this study would indicate, as a first approximation, that by devoting twenty minutes a day to instruction in each area, disadvantaged preschool children can acquire in two years' time

1. enough language learning to take them from a year or more below average up to an average level of performance,
2. two and one-half years' worth of arithmetic learning, by primary school standards,
3. slightly over one and one-half years' worth of learning in reading and spelling, by primary school standards.

In assessing the amount of achievement in reading and spelling, it must, of course, be taken into account that the normative comparison is with first-grade children who typically receive several hours a day of instruction in these areas.

3.14 Discussion

Let us return now to the concern which motivated this study in the first place. We were interested in carrying out direct instruction in a few significant areas with naive children to see how much could be accomplished with an essentially non-psychological approach--one that was developed simply out of a consideration of task requirements and the initial state of learning of the children--and to find out what problems if any remained that might require the application of principles peculiar to young children. The issue, put simply, is one of studying the task and what it requires versus studying the child and how he functions. The issue is not easily drawn, however. It must not be supposed that we are putting an arm-chair analysis of content against empirical study of children's learning. There is much to be learned about task requirements from empirical study. As an adult tries to teach live children he discovers task requirements that he could not have hit upon through speculation, if only because he himself is too far beyond the time when he first learned them and because he was too young then to know what was going on. Also, it is only as one tries to teach things to children that clear evidence of what they know and don't know emerges; test data are usually too gross and observations of children's spontaneous behavior tend to show what they can do but not what they can't.

As to the "teachability" of naive four-year-old children, there was little doubt from the first day of the experiment that the children could respond to efforts to teach them--that is, that they could assume the role of pupil as distinct from that of playmate or independent learner. We were somewhat misled in this, however, by a few of the children who apparently assumed the role of pupil as if it were a role in a play, who saw it as a matter of going through certain motions and saying certain lines on cue. They did not seem to get the idea that they were supposed to be doing a job, which was learning. This confusion of role with job is not uncommon. Little League baseball players quickly learn to play the roles of pitcher or catcher in a convincing manner, acquring many of the mannerisms of professional ball players; but only after some time and under some pressure do they learn to apply themselves to the task of getting the batter out. John Holt's book, How Children Fail (1964), is full of examples of older and vastly more sophisticated children fooling their teachers and themselves by playing the part of student while by-passing the task. And one sees many neophyte executives, military officers, engineers, and college professors who self-consciously play their roles with no apparent grasp of the accomplishments that are expected of them.

Thus the problem is not peculiar to young children. The difficulty seems to be that what is overt is easily learned while what goes on inside--what we allude to by such terms as attention and intention--is not so readily grasped. The solution proposed by Holt (1964) amounts to giving up on teaching and merely contriving to make it possible for the child to learn on his own initiative. No such drastic recourse seems called for, however. In the first place, a large majority of the children did seem to catch on readily to the task requirements of studentship. Before giving up on the others it would be well to search for ways of teaching children "how to be taught"--teaching them to recognize and cope with the problems presented rather than only to mimic the overt actions of learning. In subsequent classes we have tried to do this by introducing at early points in the program exercises that demonstrate the truth-falsity dimension of language as contrasted with the conformity-nonconformity dimension and by contriving tasks in which mere guessing or aping previously rewarded responses is dramatically unprofitable.

Arithmetic Learning

Measured by terminal test results, arithmetic was the most successfully taught of the three subjects. It is impossible to compare the starting points of the children in arithmetic, language, and reading, however, and so there is no objective basis for judging in what areas they progressed farthest from where they began. It seems reasonable to suppose that the average child's grasp of number at age four is not as far advanced over that of the disadvantaged child as his mastery of language. In studies carried out a year after the present study was begun, a group of disadvantaged children and a group of middle-class children (see Section 2.2) were taught for the same length of time with a fairly similar program. At the end of seven months the middle-class children were approximately a year ahead of the disadvantaged children on relevant ITPA language subtests, 1.3 years ahead in reading, .6 years ahead in spelling, but only .4 years ahead in arithmetic.

A cognitively-oriented psychologist might well ask whether the children, although able to perform arithmetic computations and solve a few "thought" problems, really attained an understanding of number and arithmetic. Such a question is legitimate, but it is difficult to see what criteria one might apply that would be realistic for children of this age. It would be easy enough to demonstrate, by probing with the kinds of questions that Piaget has used in studying children's concepts of number, that none of the children "really" understood number. But this is evidently also the case with advantaged kindergarten children who have been taught number concepts by a variety of methods.

In mathematics, perhaps more than in any other subject, judgments based only on what a child can do when left to cope with a problem by himself are likely to underestimate what he has learned by a large amount. A more positive approach to an assessment of children's learning in arithmetic would be that suggested by Vygotsky:

> Having found that the mental age of two children was, let us say, eight, we gave each of them harder problems than he could manage on his own and provided some slight assistance: the first step in a solution, a leading question, or some other form of help. We discovered that one child could, in co-operation, solve problems designed for twelve-year-olds, while the other could not go beyond problems intended for nine-year-olds. The discrepancy between a child's actual mental age and the level he reaches in solving problems with assistance indicates the zone of his proximal development; in our example, this zone is four for the first child and one for the second. ... Experience has shown that the child with the larger zone of proximal development will do much better in school. (Vygotsky, 1962, p. 103)

A film made by Engelmann, using children from this study at the end of their kindergarten year, gives striking evidence of the kinds of problems they could solve, with slight amounts of guidance to set them in the right direction. They could solve area of rectangle problems, side of rectangle problems, money problems involving addition, subtraction, or multiplication, factoring problems, and simultaneous equations in two unknowns. The kinds of problems they could solve without any guidance or orientation, as shown by the Wide-Range Achievement Test, put them well in advance of their years, but the "zone of proximal development" would appear to be greater even than those scores would suggest.

Even so, it can be admitted that the children did not "understand" associativity, commutativity, equivalence, the decimal base, or any of these concepts in a mature way. As Vygotsky saw, perhaps not very clearly, these are not learnings in the usual sense. They are higher-order properties of learning that come gradually to characterize the child's knowledge as he proceeds incrementally through a series of steps (no one of which decisively marks the acquisition of one or another of these understandings).

> The child is not taught the decimal system as such; he is taught to write figures, to add and to multiply, to solve problems, and out of all this some general concept of the decimal system eventually emerges. (Vygotsky, 1962, p. 102)

What Vygotsky seems to be getting at is that the child, through mastering specific concepts and operations, acquires a "working knowledge" of the decimal system, associativity, commutativity, and so on, with or without an explicit awareness of these principles, much as the child acquires a "working knowledge" of the rules of grammar, even though it may be years before he is in a position to understand them in a way the grammarian does.

Reading

Although the test results indicate that the children did make considerable progress in reading for the amount of time devoted to it (approximately 120 hours spread over the two years), it was in reading that the children exhibited the most frequent and frustrating failures to "get it" and "hold on to it." The only children who could really be said to know how to read at the end of the kindergarten year were several girls whose mothers had given them a considerable amount of help and encouragement at home. Perhaps that is the practical solution for achieving greater success in teaching reading to young disadvantaged children, enlisting parents to see that the children get enough additional practice to ensure mastery of the skill.

Our present interest, however, must be in the kinds of difficulties encountered and in the question of whether they point to cognitive peculiarities of young children that must influence reading instruction. The slower half of the children took an inordinate amount of time to learn the letter identifications and they showed a distressing tendency, even into the second year, to keep forgetting them. Visual perception seemed to have nothing to do with it. The children quickly learned to discriminate and match letters. The only child who showed visual perception anomalies (copying letters upside down, failing to match letters that were drawn slightly differently) was one of the more verbally advanced children, whose older siblings also had a history of such difficulties. He typified the "clinic cases" that have been used by remedial reading experts as a basis for generalization to reading difficulties of all kinds (Frostig, Lefever, & Wittlesey, 1963) but his problems were clearly not those of young children or disadvantaged children in general.

The second major difficulty was in learning to use letter sounds as clues to word identification. This showed up not only in the familiar inability to blend phonic elements (putting *sss-ih-t* together to form *sit*), but more dramatically in the inability to form correspondences between phonic elements and words. It is one thing to blend *sss-ih-t* as *sat* or *sick* and quite another to sound out *sss-ih-t* and then guess that the word is *run*. In the first case the child is at least using the phonic clues, though imperfectly, whereas in the second case the child shows no inkling that the phonic elements tell him anything.

The two difficulties are obviously related. The child whose mastery of the phonic identities of the letters is unstable does not have dependable phonic clues to work with. Conversely, the child who does not grasp what it is that phonic clues are able to tell him cannot be expected to appreciate the importance of remembering the letter identities. Thus we have the makings of a classic double-bind in which each of two things must be learned before the other can be learned. Such double-binds in learning tend to work themselves out in time as the child learns a little of this, a little of that, and as he acquires enough related skills from other sources to facilitate the learning of both parts. It would seem that double- or multiple-binds of this kind are very common in complex learning and that the allowing of them to be worked out over time is what lies behind the concept of readiness.

Viewed in this way, however, readiness derives not from intrinsic properties of maturation but from snarls in the sequencing of instruction. The need to "wait for readiness" disappears as the snarls are eliminated. Thus, in the case before us, we are attempting in the present year, and with encouraging early signs of success, to resolve the double-bind by teaching the children to use phonic cues in isolation from printed letters. The children are told little stories, for instance, in which key words are sounded out rather than pronounced. The children are thus provided with contextual prompts to help them identify the words. The idea is that by thus teaching the children what they can do with phonic cues, and then showing them how these cues can be supplied by printed characters, it will be easier to get across to the child that letters have stable identities that are worth-while to learn. The much greater ease with which middle-class children are usually able to learn the alphabet has no evident direct relation to language maturity. A more likely explanation is that the middle-class child learns through experience with a number of related instances that, when an adult sets out to teach him the identities of the elements of a set which he had previously treated as homogeneous, there will be some benefit from learning these identities even though he does not know what it is in advance. Engelmann has noted, for instance, that when it is pointed out to bright middle-class children that the notes in the musical scale have separate identities and that these are discernible, the children quickly learn absolute pitch without the need for intensive training.

Language

Language learning cannot be as neatly assessed as the learning of arithmetic or reading, mainly because in language it is what the child can do with what he has learned that counts whereas in arithmetic and reading one may consider certain key accomplishments as markers of the child's learning quite apart from his ability to apply arithmetic or understand various written material. As to the general practicality of teaching new concepts and syntactical forms through the methods of demonstration, drill, and application that we employed, we can say that the methods are highly reliable. But, then, so are the informal methods of language teaching that go on in the home highly reliable, in that it can be assumed with great confidence that in time children will master the complete grammar and everyday vocabulary of their parents. It is all a matter of relative efficiency.

In judging efficiency we must rely on the rather gross evidence provided by test data. The rapid progress made by the subjects in this study in learning grammatical inflections (as shown by scores on the Auditory-Vocal Automatic subtest of the ITPA) during the first year, and the virtual cessation of this progress when direct instruction was abandoned in the second year, testify to the efficiency of direct as against informal teaching. For the children were certainly exposed to models of use of grammatical inflections during the second year, perhaps more richly than during the first year, and they were also subjected informally to corrections of their own speech; but this proved to be inadequate for maintaining normal progress.

In light of our more fundamental interest in teaching the children to use language to arrive at valid assertions about reality, we must note that in spite of their improved performance on tests of verbal reasoning, the children impressed us as falling short of the level of sophistication that bright middle-class children show in the logical use of language. This, we fear, will remain their major handicap in academic days to come. In a nutshell, they didn't learn to ask themselves the right questions and use the answers to these questions to ask more questions so as to arrive eventually at a solution or explanation. Given help, in the form of someone who asked the questions for them, they could accomplish dramatic feats, solving scientific problems, arriving at exclusive definitions, and so on. But they didn't learn how to pose questions themselves, except in the highly structured framework of arithmetic.

This is not to say that the children did not ask questions. They asked the kinds of "curiosity" questions that on superficial observation would suggest an adequate degree of inquisitiveness, "exploratory drive," or what have you. But these are the kinds of what and why questions that are elicited merely by exposing the child to a novel phenomenon. These are the kinds of questions one supposes a chimpanzee would ask if it knew how. Some progress was made during the second year by teaching the children to ask what unfamiliar words in instructions meant and to plug these meanings back into the instructions so as to render them intelligible ("Touch your cranium." "What's a cranium?" "It's your head bone.") and by introducing the children to the game of "Twenty Questions." A comprehensive attack on the problem of teaching self-questioning remains to be devised, however. At this time we can only register the surmise that children need to become quite surefooted in dealing logically with questions that are put to them before they can achieve much success at or even appreciate the value of asking themselves questions. For children who start with as little mastery of language and as little conception of truth and falsity as the children we worked with, therefore, this may well be the longest row they have to hoe and one which must be continued with some care beyond the two years of preschool education.

3.15 Conclusion

The preceding discussion dealt largely with apparent short-comings of the experimental instructional program. It should be noted that these short-comings could not have been detected if the program had not been so successful in what it did do. It is impossible to tell whether a beginning golfer has a slice until he has learned to hit the ball. For that reason it is impossible to judge whether children trained in less achievement-oriented programs would have similar problems. An important question to consider, however, is whether the shortcomings that remain can be dealt with by pedagogical means--by adding elements to the curriculum that had been neglected, by removing or altering sources of mislearning, by rearranging sequences, clarifying instructions, improving methods of presentation and management, etc.--or whether the difficulties lie outside the scope of pedagogy and require accommodation to maturational or psychological factors of a special sort. No general answer can be given to this question. It can be answered, by demonstration, with each difficulty as it appears. We have, however, attempted to show in the preceding discussion, that for each of the variety of difficulties considered, there exist substantial resources for dealing with them within the instructional system that we have developed. These problems can, in other words, be conceived of as learning problems that can be subjected to the same kind of analysis and attacked by the same pedagogical methods as the problems that were successfully dealt with by instruction.

In short, we believe this study indicates that approaching early learning as a straight-forward instructional problem, as later school learning is approached, is a productive and progressive way not only of developing effective educational programs for young children but also of identifying the constituents of those amorphous categories that now go by the names of "intellectual development" and "readiness."

3.16 Summary

Fifteen four-year-old disadvantaged children were chosen on the basis of older children in the same households having educational difficulties that were attributed by their teachers to cultural deprivation. For two years they were given a special educational program that featured one hour a day of direct verbal instruction in language, arithmetic, and reading. The remainder of the program (an additional hour the first year, an hour and half the second year) consisted of less structured activities designed to reinforce the effects of direct instruction.

Relevant test data included a test of ability to use grammatical inflections (the Auditory-Vocal Automatic subtest of the Illinois Test of Psycholinguistic Abilities), a verbal analogies test (the Auditory-Vocal Association subtest of the ITPA), a test of expressive language (ITPA Vocal Encoding), a vocabulary test (ITPA Auditory Decoding), general intelligence (Stanford-Binet MA), and achievement in reading, arithmetic, and spelling (Wide-Range Achievement Test). On first testing the children averaged from one-half to one and one-half years below average on all language tests. By the end of the second year their scores had risen to approximately average, except that on the grammar test their scores rose from far below average to within two months of average the first year, but then direct instruction in grammatical usage was suspended and they gained only four months the next year. As a result they ended with scores again well below average on this test. Mental age rose from about six months below average (the first testing occurring six weeks after school started) to about four months above average. Terminal achievement averaged at the 1.5 grade level in reading , 2.6 in arithmetic, and 1.7, in spelling.

3.2 Direct Verbal Instruction Contrasted with Montessori Methods in the Teaching of Normal Four-Year-Old Children

3.21 Purpose

An experimental comparison between the methods of direct verbal instruction described in section 3.1 and Montessori methods of preschool education has several points of interest. The two approaches share certain common goals that are not shared by traditional approaches to preschool education. Both are concerned with teaching, with getting children to know things they don't already know. Both emphasize work instead of play. Both follow the principle of presenting sequenced tasks of increasingly more stringent criteria, so that the child has the experience of encountering and overcoming failure rather than being spared it altogether. More specifically, both attempt to teach mathematical concepts, reading, and spelling to preschool children.

In spite of this communality of goals and interests, the two approaches are diametrically opposed on almost every major point of method. The over-riding difference, of course, is that direct verbal instruction works mainly through language, both in the teacher's presentation and the children's responses, whereas Montessori methods are far to the opposite extreme, relying on the child's physical interaction with inanimate objects to accomplish most of the learning, verbal behavior by both teacher and child being minimized. In addition, Montessori places great emphasis on the child's freedom to select and pace his own activities, whereas in direct verbal instruction the selection and pacing are controlled by the teacher. Children usually work separately in Montessori schools, but in groups under direct verbal instruction. The atmosphere strived for in Montessori schools is one of library-like calmness and concentration, whereas with direct verbal instruction it is more one of gymnasium-like strenuous exertion.

The present experimental comparison was concerned with two issues: 1) whether the contrasting verbal and non-verbal emphases of the two methods would be reflected in differential gains on verbal and non-verbal performance measures and 2) whether direct, teacher controlled and paced instruction in basic academic subjects would produce different achievement results than the less direct and forceful instruction of the Montessori method. With respect to the first issue, it was expected that the Montessori group would show up better on non-verbal tests, the Academic Preschool group on verbal tests, except for the grammar and vocabulary subtests, where no difference was predicted. Since almost all of the children came from relatively privileged homes with highly educated parents, it was expected that those verbal measures that reflected fundamental language skills would not be materially affected by differences in school treatment. With respect to the second issue, it was expected that direct verbal instruction would lead to higher levels of mean performance on achievement tests, but that the variance of scores would be larger and negatively skewed for Montessori-trained children, indicating that most scored below the mean of the direct verbal instruction group, but that some scored as high as or higher than the best of the direct verbal instruction group.

3.22 Method and procedure

Subjects.

Subjects for the Montessori group were all the children enrolled in a local Montessori school who would be of kindergarten age the following academic year. This meant that they were between the ages of 3-10 and 4-10 at the time the experiment began. Nineteen children were so identified, of whom 2 were later excluded for reasons to be given. The school was a licensed Montessori school that had been in existence for two years, using trained Montessori directresses. Children in the Montessori school were largely from upper-middle-class homes with college educated parents.

The recruitment of a comparable group of children to receive instruction under the direct verbal instruction method was carried out somewhat unconventionally because of the special nature of the Montessori group. It would have been easy enough to have recruited a group from the same socio-economic backgrounds as the Montessori children, but there was reason to suppose that the parents of the Montessori children differed from the average in ways not indicated by their occupational and social status. The Montessori society was still young in the community and many of the parents had taken part in the original movement to start such a school. The writer attended one of the planning meetings and observed the parents to be exceedingly zealous over their children's intellectual advancement and convinced that regular nursery schools were inadequate because they "didn't teach anything." It could be expected, therefore, that such parents would take an even more active role in their children's early education than is common for people of comparable station.

Accordingly, recruitment was carried out by telephoning parents of children in the Montessori school and securing from them names of parents whom they knew to be interested in the Montessori school and desirous of having their four-year-old child in it, but who were unable to enroll him either because of the Montessori school's policy of not admitting children above the age of three years six months or because of the financial burden (tuition in the Montessori school being high by local standards). Some names were also obtained from the waiting list for admission to the Montessori school. Through this means 19 subjects were recruited, of whom one was subsequently lost by leaving the community.

One important difference, however, that could not be eliminated so long as intact Montessori classes were used, was the fact that the Montessori children had already had a year or more of schooling in the Montessori school at the time the present study began, whereas none of the children in the direct verbal instruction group had been in school previously.

Educational Treatment

Except for the testing to be described, the children in the Montessori school were experimentally untouched. The four-year-olds being studied were in two different classes, mixed in with a number of younger children and a few older ones, according to the Montessori principle of avoiding age segregation. The school day was three hours.

The class to be taught by direct verbal instruction methods was started in mid-October, under the name of the Academic Preschool.

It followed approximately the same schedule of days as the Montessori school, but the school day was only two hours. Initially the same basic program was followed as with the disadvantaged children the year previously--three teachers each teaching a different subject in 20-minute sections to groups of children stratified according to level of attainment. The teachers were all new to the program. One had previously taught slow-learners in elementary school and another had taught high school English. The major departures from the plan used with the disadvantaged children were as follows:

1. Since the children in the present study were much more sophisticated in language usage than the disadvantaged children, the basic language program was presented only in the most summary form. Most of the time was devoted to the material on if-then reasoning from the advanced language program and to science topics that also involved verbal reasoning problems; e.g., problems involving opposing forces, problems involving deductions from the geological record.

2. A revised reading curriculum was used, employing the Initial Teaching Alphabet. The new curriculum included early introduction of rhyming and first-letter phonics to distinguish rhyming words, early introduction of sentence reading with a small set of sight words, and very systematic drill on spelling patterns.

3. Initially the arithmetic curriculum was modified to accommodate early introduction of negative numbers. However, teaching difficulties required an eventual retreat back to the arithmetic program as used with the disadvantaged children.

4. Printing (in i.t.a.) was taught informally but regularly as one of the between-class activities.

5. After four months a fourth subject was added, experimental science, in which a number of short units centering around demonstrations and problems in rule-application were tried out. These included units on plant physiology, weather, and geology.

6. At this same time the class was divided into four groups for reading instruction instead of the usual three because four children had gotten too far ahead of the others in reading to profit from the same instruction as the others were receiving.

7. The original schedule of spreading the instructional sessions over two hours by sandwiching periods of semi-structured activity between them was gradually modified until the instructional periods were run one immediately after the other, the less structured activities being all shoved to the end of the day.

Testing

The Illinois Test of Psycholinguistic Abilities (ITPA) was administered both as a pretest and as a posttest. The first administration was carried out approximately six weeks after the beginning of the Academic Preschool. One child in the Montessori school refused to be tested on this and later occasions and was therefore dropped from the sample. Another Montessori child obtained a standard score on the total test that was 1.2 standard deviations below that of any other child in either group (a gap four times greater than any other in the score distribution) and therefore was excluded as a case lying outside the range to which results of the study might be generalizable. The ITPA was readministered to both groups approximately six months later.

At the latter time an individual achievement test was also administered, the Wide-Range Achievement Test, which provides grade-level scores for reading, arithmetic, and spelling that range down as far as the beginning kindergarten level (0.1). The reading test consists entirely of sight vocabulary words, except for some alphabet recognition items at the lowest levels. Because both the Montessori children and those in the Academic Preschool had been taught to identify letters by sound rather than name (although many of them had learned the names at home), the scoring policy was followed (which is admissible under the rules for administering the test) of giving credit for all the letter identifications providing the child recognized one or more words. The words were presented on the standard test forms to both groups--that is, in traditional orthography, with lower-case letters--even though the Academic Preschool children had been taught with the Initial Teaching Alphabet. The arithmetic test consisted of 15 orally administered items and had a ceiling score of 2.0 grade level. The spelling test consisted of individual words graded in difficulty and was scored conventionally. Thus, words spelled correctly in the Initial Teaching Alphabet were not counted as correct, unless the spelling happened to conform exactly to conventional orthography.

3.23 Results

Pretest results

Table 3.23.1 gives a comparison of the pretest scores on the ITPA for the Montessori and Academic Preschool groups. The data are in standard scores fitted to the nearest month of chronological age by means of linear interpolation in the norm tables. Separate *t*-tests were run on each ITPA subtest as well as on total score. None of the differences was significant even at the .10 level, even though with 10 comparisons one of the differences could have been expected to reach that level by chance. Thus the groups may be regarded as not differing significantly, either in overall level of test performance or in particular subtests. The testers inadvertently missed two children in the Academic Preschool and the oversight was not made known until it was too late to obtain usable pretest scores. These two children are left out of further analysis of ITPA scores; however, their scores on the posttest would indicate that they were slightly below the average of their group, so that their inclusion in the pretest comparison could have been expected to diminish rather than augment the overall difference.

ITPA posttest results

Analysis of covariance was used to remove some of the experimental error due to within-group differences and (presumably) random between-group differences shown on the pretest. A single repeated-measurements design using corresponding subtest scores in the pretest as the covariate, was intended, but preliminary tests showed significant heterogeneity of regression among the ITPA subtests ($p < .01$), thus making it unreasonable to base all adjustments on the same regression coefficient. Accordingly, a separate analysis of covariance was run for each subtest as well as for total test score. Tests on within-groups regression coefficients showed

Table 3.23.1 - ITPA pretest standard scores for direct instruction and Montessori groups

Test	Direct Instruction(N=16) Mean	Direct Instruction(N=16) S.D.	Montessori(N=17) Mean	Montessori(N=17) S.D.	t*
1. Aud. Decoding	1.01	.81	.75	.66	.99
2. Vis. Decoding	.45	.68	.78	.62	1.41
3. Aud-Voc. Association	1.33	.69	1.40	.69	.29
4. Vis-Mot. Association	.11	.68	.17	.74	.24
5. Voc. Encoding	1.01	.90	.65	.77	1.19
6. Mot. Encoding	.20	.70	.41	.84	.76
7. Aud-Voc. Automatic	.84	.86	.61	.53	.90
8. Aud-Voc. Sequential	.98	.92	.78	.95	.60
9. Vis-Mot. Sequential	.75	.74	.83	.49	.37
Total Score	1.33	.61	1.28	.67	.22

*_t_ of 1.70 required for significance at .10 level

no significant heterogeneity at the .10 level on any test. Table 3.23.2 reports the posttest means and adjusted means, along with F values for the analyses of covariance.

On total test score, the groups differed hardly at all. In fact, there was practically no change for either group from the pretest scores--a gain of .07 standard score points for the Academic Preschool group and a loss of .11 for the Montessori group. On two of the subtests, however, there were significant adjusted posttest differences. One is a difference of half a standard score unit in favor of the Academic Preschool group on Auditory-Vocal Association and the other is a difference of equal magnitude in

Table 3.23.2 - Adjusted and unadjusted posttest means on ITPA

	Group					
	Direct Inst. (N=16)		Montessori (N=17)		Adjusted	
Test	Unadjusted Mean	Adjusted Mean	Unadjusted Mean	Adjusted Mean	Mean Differ.	p*
1. Aud. Decoding	.81	.78	.72	.74	.04	n.s.
2. Vis. Decoding	.23	.25	.79	.78	-.53	<.05
3. Aud-Voc. Association	1.64	1.67	1.16	1.13	.53	<.01
4. Vis-Mot. Association	.96	.97	.51	.49	.48	<.25
5. Voc. Encoding	.93	.89	.36	.40	.49	<.25
6. Mot. Encoding	.12	.18	.26	.20	-.03	n.s.
7. Aud-Voc. Automatic	.87	.81	.73	.78	.03	n.s.
8. Aud-Voc. Sequential	.90	.83	.82	.89	-.06	n.s.
9. Vis-Mot. Sequential	-.39	-.37	-.10	-.13	-.24	n.s.
Total Score	1.40	1.38	1.17	1.19	.19	n.s.

* P based on analysis of covariance of posttest scores, controlling for corresponding pretest scores (d.f. = 1,30).

favor of the Montessori group on Visual Decoding. Auditory-Vocal Association is a verbal analogies test. Visual Decoding is a test in which the subject is briefly shown a picture of an object and then must locate an object of the same identity in another picture. Two other differences approached significance, both being on the order of half a standard score unit. Both differences were in favor of the Academic Preschool group. One was on Vocal Encoding, a test that requires the subject to "tell all about" objects that are given him, the score being the number of different attributes he names. The other was Visual-Motor Association, a picture analogies test.

An overall comparison of performance on verbal and non-verbal subtests reveals the following: on verbal subtests (1, 3, 5, 7, 8) the Montessori group declined an average of .08 standard score points from pretest to posttest, while the Academic Preschool group showed no change. On the non-verbal subtests (2, 4, 6, 9) the Montessori group showed a decline of .18 while the Academic Preschool group showed a decline of .04. The differences between groups from this perspective are minute and contrary to the expectation that the contrasting verbal and non-verbal treatments given the two groups would be reflected positively in ITPA subtest changes.

Another comparison was run which, because of its glaringly *a posteriori* nature, must be taken with some salt. However, it represents a way of looking at ITPA profiles that may be more meaningful for normal children than the diagnostic categorizations developed for exceptional children. Subtest scores are assigned weights according to the amount of information processing as contrasted with rote or automatic reproduction that they seem to call for. The two analogies tests, Auditory-Vocal Association and Visual-Motor Association, are assigned weights of +4. Auditory Decoding (a vocabulary test requiring a judgment as to whether a given statment can be true or not) and Vocal Encoding (described above) are assigned weights of +2. Visual Decoding (described above), Auditory-Vocal Automatic (grammatical usage), and Motor Encoding (showing by pantomime what to do with pictured objects) are weighted -2; and the two immediate memory tests, Auditory-Vocal Sequential and Visual-Motor Sequential, are weighted -3.

Thus the composite score has an expected value of zero, postive scores indicating a tendency to score higher on tests requiring information processing than on more automatic tests, negative scores indicating the opposite tendency.

Mean scores for the two preschool groups on this composite were postive for both pretest and posttest. On the pretest the Academic Preschool group obtained a mean score of 1.48 and the Montessori group a mean of .53. The difference was not significant (t=.55, d.f.=31). On the posttest the mean for the Academic Preschool group rose to 8.93 and the mean for the Montessori group rose to 2.70. An analysis of covariance (controlling for pretest score) showed this posttest difference to be significant at the .02 level (F= 7.22, d.f.=1,30).

Achievement Test Results

Table 3.23.3 presents means, standard deviations, and ranges for achievement test scores. As predicted, the Academic Preschool children were significantly superior in all three areas. The Montessori group showed greater variance in scores on two of the three tests. The difference was significant (by F test) at the .01 level in arithmetic and at the .05 level in spelling. On reading, however, the variance was insignificantly greater for the Academic Preschool group. The range was curtailed in the arithmetic test by a ceiling of 2.0 grade level. Two children in the Academic Preschool group and one in the Montessori group reached this ceiling.

Table 3.23.3 - Means, standard deviations, and ranges of scores on Wide-Range Achievement Test (grade-level scores)

	Group							
	Direct Instruction(N=18)			Montessori (N=17)				
	Mean	S.D.	Range	Mean	S.D.	Range	t	p
Reading	2.43	.68	1.4-3.7	1.05	.62	.0-2.6	6.30	< .005
Arithmetic	1.46	.22	1.2-2.0	1.19	.40	.3-2.0	2.45	< .025
Spelling	1.72	.32	1.0-2.2	1.25	.51	.8-2.5	3.62	< .005

Contrary to expectations, the Montessori group did not show marked negative skewness in scores, except in spelling. As a crude comparison of skewness, the number of children in each group scoring above and below their respective group means was compared by a chi-square test. In reading the Academic Preschool group was evenly divided and the Montessori group was divided 9 above and 8 below--clearly not a significant difference, with no skewness in either group. On arithmetic, scores for the Academic Preschool group were negatively skewed--5 above and 13 below while for the Montessori group they were positively skewed--10 above and 7 below. The difference fell short of significance at the .10 level, however. On spelling the results were reversed--11 above and 7 below for the academic Preschool group and 5 above and 12 below for the Montessori group. Again the difference was not significant.

3.24 Discussion

The two major questions in this study were 1) whether the differences between the highly verbal training given in the Academic Preschool and the highly non-verbal training given in the Montessori school would be reflected in differential gains on verbal and non-verbal ITPA subtests and 2) whether the differences in treatment would result in differences in academic achievement.

The ITPA differences obtained seem to be better accounted for on the basis of the amount and kind of information processing required of children in the two schools rather than on the basis of modality. Thus, of the three tests on which the Academic Preschool children most exceeded the Montessori children, one is non-verbal, but all have in common the following elementary task requirement: the child, on being presented a stimulus (either verbal or non-verbal), must abstract some characteristic or characteristics from it according to criteria of task relevance. The one subtest on which the Montessori children did significantly better, Visual Decoding, was superficially very similar to the Visual-Motor Association subtest, on which the direct instruction group was superior. On Visual Decoding, however, the child does not need to abstract any attribute; he need only recognize the object shown and then locate another object of the same identity (whereas on Visual-Motor Association he must locate an object that "goes with" the stimulus object--i.e., that is related to it by use, class membership, etc.).

One of the peculiarities of Montessori instruction, besides its non-verbal character, is that most of the abstraction is done for the child rather than by him. If he is to attend to length, color, or pitch, then all the attributes of the stimulus objects are held constant except length, color, or tone. In our program the same principle is used in introducing new concepts, but instruction always goes on from there to ensure that the new dimension, attribute, or class is distinguished from others. Thus the child is eventually taught to deal with abstractions under normal conditions. Informal nursery school teaching usually presents normal ("noisy") conditions from the beginning, thus making abstraction difficult; but it seems as if the Montessori method may go too far the other way in isolating concepts. In any event, it appears from the present results that Montessori-trained children do not do better on non-verbal tasks *per se*, but only on non-verbal tasks that have a minimum of conceptual content.

The differences in achievement in reading, arithmetic, and spelling were sizable, although the Montessori children did not do badly for four-year-olds, scoring at the first-grade level or above on every test. By comparison, however, in another study carried out simultaneously with this one and covering the same time-span, disadvantaged four-year-olds given the direct instruction curriculum earned scores almost identical with those of the considerably more privileged Montessori group, while disadvantaged children given a typical nursery school program earned scores at about the .5 grade level.

Only one child in the Montessori group could be said to know how to read, scoring above the 2.0 level on both reading and spelling. Thirteen of 18 children in the Academic Preschool scored at the second grade level or above in reading, and four did so in spelling. (The discrepancy between reading and spelling performance being most reasonably attributed to use of the i.t.a. alphabet in instruction and traditional orthography in testing).

One possible objection to the results of this study is that the Academic Preschool children received a type of training that merely disposed them to do better on tests generally--to be more acquiescent in the testing situation and to try harder, perhaps. Thus the Montessori children may have known more than they revealed. This objection cannot be ruled out. To be sure, the two groups did equally well on the ITPA pretest and posttest, but the Academic Preschool children had already had six weeks of schooling at the time of the pretest, so that they may already have acquired some of these favorable test-taking behaviors. From this it would have to follow, however, that the Montessori children were actually better in ITPA abilities from the beginning. If one grants this objection, therefore it makes the achievement differences appear greater even than the scores indicate.

Comparing the present study with the earlier one on disadvantaged children, two differences emerge. One is that for children from language-rich environments, direct verbal instruction has little impact on language abilities, whereas it has a considerable impact for disadvantaged children. A higher-level language program (one that dealt with more advanced grammatical constructions, etc.) might have an impact, but it is questionable whether it would be a profitable use of instructional time. The other difference was in reading, the middle-class children making far greater progress than the disadvantaged children in the same length of time. This issue was discussed in the preceding report. The implication for educating middle-class children, however, is that reading seems to come easily enough for four-year-olds that there is little reason to delay it another two years.

3.25 Summary

A direct verbal instruction program of seven months' duration was given to a group of 18 four-year-olds, chosen to be comparable with an equal-sized group of four-year-olds who were undergoing their second year of education in a Montessori school. The Illinois Test of Psycholinguistic Abilities was administered as a pretest and posttest, and the Wide-Range Achievement Test as a posttest only. On the pretest, administered six weeks after the beginning of instruction for the direct instruction group, there was no significant standard-score difference at the .10 level either in total score or subtest scores. Total standard scores on the ITPA remained nearly constant for both groups from pretest to posttest. Using pretest score as a covariate, however, there was a significant posttest difference in favor of the direct instruction group on the Auditory-Vocal Association subtest and in favor of the Montessori group on the Visual Decoding subtest. There was, however, no general interactive tendency; instead of the Montessori group doing relatively better on the non-verbal subtests and the direct verbal instruction group doing better on the verbal subtests, the direct instruction group did insignificantly better on both types. The only discernible pattern difference seemed to do with amount of information processing required rather than with modality: the direct instruction group did relatively better on tests involving abstraction whereas the Montessori group did relatively better on tests of simple recognition or memory. An *a posteriori* analysis, using composite scores weighted on this dimension, yielded a significant difference.

On the Wide-Range Achievement Test the direct instruction group significantly surpassed the Montessori group in all areas. Mean grade-level scores for the direct instruction group and the Montessori group, respectively, were 1.46 and 1.19 on arithmetic, 3.43 and 1.01 on reading, and 1.72 and 1.25 on spelling.

3.3 Implications for Future Research in Early Education

The research described in this report has followed a familiar sequence in educational program development--from exploratory studies to the formulation of a coherent program and then to the testing of its feasibility and short-term effects. The next steps, if this familiar sequence were to be pursued, would be to carry on follow-up studies to observe effects of longer term and to test the program in different places on different types of children, meanwhile continuing to refine the program as new data or complaints came in. These are all important steps and we are, in fact, trying to pursue them, as are numerous other people who are developing early education programs.

The only trouble with this sequence of program development is that development has almost no part in it. After a program has passed its initial tests of feasibility and effectiveness, it is pretty well set and any changes brought about through more comprehensive trials and tests are likely to be of a trivial sort (unless of course the results are disastrous and the program is scrapped--but that is not development either).

With projects of a more technologically advanced sort, as in electronics, development can usually proceed in an orderly fashion even without comparative testing, although comparative data are frequently helpful. Color television sets get better and more economical year by year without the need for extensive evaluative research, because it is possible to recognize an improvement before it gets out of the workshop. The dimensions of performance for color television sets are well enough defined that it is evident without comparing one set with another in what ways each could be better, and so progress depends primarily upon the continuous production of feasible new ideas.

In education progressive development is not so simple. New ideas are relatively easy to come by (a fact little appreciated by arm-chair innovators who think it a great thing that some enterprising nursery school teacher is experimenting with teaching children to play chess). The problem is to find out whether any innovation is a step forward or backward. To evaluate every new idea under adequately controlled conditions is out of the question. A group of uninhibited graduate students could produce enough ideas in one afternoon to keep educational research busy for three hundred years. Consequently, program development is guided largely by force of personal conviction, and evaluative research is mainly of use in helping consumers decide whose personal convictions to buy.

Early education is virtually driven to comparative research methods by the absence of well-defined performance criteria. These criteria cannot simply be created by fiat or consensus. Let half a dozen television engineers examine the same TV set and they would probably agree on the ways in which it could be improved. In fact, they could probably agree without having to examine it, assuming that there is room for improvement on all performance dimensions. How these improvements could be brought about is another matter, but not a matter for polemical dispute. It is a matter for research directed to the solution of technical problems. Let half a dozen educators examine an on-going preschool program, however, and one could expect many contrary opinions about what kinds of improvements were needed. An innovation that constituted an improvement to one person would be considered irrelevant or retrograde by another. In educational program development comparative objective evaluation is a poor but inevitable substitute for the kind of guidance that would come from well-defined performance criteria. If a given program yields poor test results in comparison with standard or control-group norms, this at least indicates in a general sort of way where improvement is needed in order to bring the program up to the performance level of competing programs. But norms represent a spurious standard. Although a program might compare favorably with all others on a certain test criterion, it could well be that all existing programs are seriously deficient on this criterion. This is sometimes brought to light when comparison is shifted to a different norm group. Thus in a recently completed international study of achievement in advanced secondary-school mathematics (Hartung and Foshay, at this time noted only in newspaper reports), it appears that U.S. students as a whole perform well below international norms. Judging a mathematics project by U.S. achievement norms, therefore, or by comparison with other existing U.S. programs, would give a spurious impression of sufficiency.

As a supplement to normative testing, one can carry out continual testing of specific achievements (this, it should be noted, is only possible in the unusual preschool program where achievement objectives are rather precisely itemized). Such testing has at least the advantage of showing whether children are learning what is expected of them, but it does not permit any judgment as to whether they are learning it as efficiently or thoroughly as might be possible. On the assumption that some incidence of failure is inevitable, the program developer would like to know whether the actual incidence he observes is irreducible. On the other side, if it takes two weeks for children to attain a certain criterion, the developer would like to know whether this is close to the maximum limit of efficiency possible, so that efforts at improvement would better be directed elsewhere, or whether there may be a way of accomplishing the same thing in a tenth the time, in which case it would be well worthwhile to search for a better way.

We suffer, in other words, from an almost complete ignorance of limits and must depend on empirical judgments which may well be far off the mark. By contrast, workers in technologically more advanced fields can often derive limits from physical laws (limits of efficiency for heat engines, limits on the amount of information that can be transmitted through a channel, etc.). There is at present no alternative to empirical estimation of limits in education, but there is virtually no effort being made to pursue this source of information.

A minimum requirement for obtaining empirical estimates of limits is that identical achievement objectives be pursued by different means. Instead, comparative studies in education almost invariably do one of two things: they either compare approaches that differ in both methods and objectives (the usual practice in curriculum and program evaluations, as illustrated in Section 3.2) or else they compare sub-optimal variations on the same approach (the usual factorial study, as illustrated roughly in Section 2.4). Neither type of study provides even a crude first approximation to limits of efficiency and level of attainment on specific criteria. A better type of study can be easily envisaged. In it certain specific learning objectives are set--definite concepts, skills, or behaviors to be induced, with explicit, objective criteria of success. Then two or more program developers, following different approaches, design and test methods of attaining these objectives. They do not stop at the first test, however, but continue to develop their methods, pirating freely from one another as they see fit, until they reach points of diminishing returns or until the distinctions between their approaches have been dissipated.

Although we are accustomed, from the results of global end-of-term comparisons, to expect small differences between contrasting educational approaches, it is entirely possible that on specific tasks the productivity of two methods could differ by an order of magnitude. The informational value of such differences is obvious, but even where the differences were small, there would be substantial value in being able to observe what difficulties were common to all approaches and what ones were so-to-speak manufactured by the method.

In this way progress could be made toward the definition of technical problems or difficulties, as discussed in Section 1.0. Such studies would yield information on a day-to-day basis rather than on a year-to-year basis, as is the case with typical program comparisons.

If the advantages of this type of study are so obvious, one might ask why such studies are not already being done. One reason seems to be that educators can seldom agree on the most desirable learning objectives. There is, however, no reason why they should have to agree. It is only necessary for them to agree that for research purposes certain tasks would be of mutual interest. Presumably what was learned could be transferred later to work on those objectives which each program developer considered most worthy.

Another, more stubborn obstacle is that early childhood educators do not for the most part think in terms of specific learning objectives at all. When educational objectives are framed only in terms of the fostering of various traits, attitudes, or aspects of development, there is no profitable way of aligning programs step-by-step and part-by-part. Usually there are no steps in the curriculum and although the curriculum may have parts, every part is related to a number of different objectives. When this is the case, there is not much to compare between programs other than their ultimate effects. In the resume of the exploratory studies (Sect. 2.7), we argued that this global view of early education and intellectual development is educationally meaningless. Here we may add the argument that, although worthwhile educational programs might be conceived from this orientation, they lack the researchability that is necessary for them to develop beyond their original formulations. Globally conceived educational programs can be defended, and there is little doubt that they will survive, if only by force of tradition; but they have little to offer as vehicles for educational research and development.

Research in early education would do better to concern itself with specific learning tasks. Even if, in the end, it were decided that less structured, globally conceived programs were more desirable for some reason, they would gain more from this kind of research than from research carried on within their own amorphous structures and based upon their sweeping principles. In this regard, it is hoped that the studies reported here will at least give a start to the defining of learning tasks and task-related problems at the preschool level.

References

1. Atkin, J. M. Some evaluation problems in a course content improvement project. *J. Res. Science Teaching*, 1963, 1, 129-132.

2. Ausubel, D. P. *The psychology of meaningful verbal learning*. New York: Grune and Stratton, 1963.

3. Ausubel, D. P. How reversible are the cognitive and motivational effects of cultural deprivation? Implications for teaching the culturally deprived child. *Urban Education*, 1964, *1*, 16-38.

4. Bender, M. L. The value of pictures in machine-teaching of sight vocabulary to preschool children. Unpublished master's thesis, University of Illinois, 1965.

5. Bender, M. L. Control of attention and reduction of trial-and-error responses in teaching machines. Unpublished doctoral dissertation, University of Illinois, 1966.

6. Bloomfield, L., & Barnhard, C. L. *Let's read*. Detroit: Wayne University Press, 1961.

7. Boyd, W. *The Emile of Jean Jacques Rousseau*. New York: Teacher's College, Columbia University, 1962.

8. Braine, M. D. S. Piaget on reasoning: a methodological critique and alternative proposals. *Monogr. Soc. Res. Child Developm.*, 1962, 27, no. 2, 41-61.

9. Brison, D. W. Acquisition of conservation of substance in a group situation. Unpublished doctoral dissertation, University of Illinois, 1965.

10. Bruner, J. S. *The process of education*. Cambridge, Mass.: Harvard University Press, 1960.

11. Bruner, J. S. The course of cognitive growth. *Amer. J. Psychol.*, 1961, 19, 1-16.

12. Carroll, J. B. The analysis of reading instruction: perspectives from psychology and linguistics. In E. R. Hilgard (Ed.), *Theories of learning and instruction, Yearb. nat. Soc. Stud. Educ.*, 1964, 63 (1), 336-353.

13. Charlesworth, W. R. Instigation and maintenance of curiosity behavior as a function of surprise versus novel and familiar stimuli. *Child Developm.*, 1964, 35, 1160-1186.

14. Cronbach, L. J. *Educational psychology* (2nd ed.). New York: Harcourt, Brace, and World, 1963.

15. Curtis, S. J., & Boultwood, M. E. A. *A short history of educational ideas*. London: University Tutorial Press, 1956.

16. Dolch, E. W., & Bloomster, M. Phonic readiness. Elem. Sch. J., 1937, 38, 201-205.

17. Doman, G., and others. You can teach your baby to read. Ladies Home J., May, 1963, 80, 62-63+.

18. Dreikurs, R. The ABC's of guiding the child. Chicago, Comm. Studying Mental Health in the Public Schools, Sears-Roebuck Foundation, undated.

19. Flavell, J. H. The developmental psychology of Jean Piaget. Princeton, N. J.: Van Nostrand, 1963.

20. Fowler, W. Teaching a two-year-old to read: an experiment in early childhood learning. Genet. Psychol. Monogr., 1962, 66, 181-283.

21. Friedlander, B. Z., & Dixon, J. The effect of repetition, variation, and randomness on preschool children's free-play performance for perceptual reinforcement. Unpublished report, Western Reserve University, undated.

22. Frostig, M., Lefever, D. W., & Whittlesey, J. R. B. Visual perception in the brain-injured child. Amer. J. Orthopsychiat., 1963, 33, 665-671.

23. Gagne, R. M., and others . Factors in acquiring knowledge of a mathematical task. Psychol. Monogr., 1962, 76, No. 7 (Whole No. 518).

24. Hiveley, W. Some guidelines for evaluating programs of instruction in mathematics. Unpublished paper, University of Minnesota, undated.

25. Holt, J. How children fail. New York: Pitman, 1964.

26. Hunt, J. McV. Intelligence and experience. New York: The Ronald Press Co., 1961.

27. Hunt, J. McV. Motivation inherent in information processing and action. In O. S. Harvey (Ed.), Motivation and social interaction: the cognitive determinants. New York: Ronald Press, 1963.

28. Inhelder, B., & Piaget, J. The growth of logical thinking from childhood to adolescence. New York: Basic Books, 1958.

29. Kessen, W. "Stage" and "structure" in the study of children. In W. Kessen & C. Kuhlman (Eds.), Thought in the young child. Monogr. Soc. Res. Child. Developm., 1962, 27, Nr. 2 (Whole No. 83), 65-86.

30. Leonard, G. B. What your child can teach his teacher. Look, Dec., 1966, 29-30.

31. Lovell, K. The concept of substance. In P. Mussen, J. Conger, J. Kagan (Eds.), Readings on child development and personality. New York: Harper and Row, 1965, 290-297.

32. Lowenfeld, V. Creative and mental growth. New York: Macmillan, 1949.

33. Montessori, M. The Montessori method: scientific pedagogy as applied to child education in the "children's homes." New York: F. A. Stokes Co., 1912.

34. Moore, O. K. Autotelic responsive environments and exceptional children. Hamden, Conn.: Responsive Environments Found., 1963.

35. Piaget, J. The moral judgment of the child. New York: Harcourt, Brace, and World, 1932.

36. Piaget, J. The psychology of intelligence. (Trans. by M. Piercz and D. E. Berlyne) Paterson, N. J.: Littlefield, Adams and Co., 1960.

37. Piaget, J. Development and learning. In R. E. Ripple and V. N. Rockcastle (Eds.), Piaget rediscovered. Ithaca, N. Y.: Cornell University, School of Education, 1964, pp. 7-20.

38. Pines, M. A pressure cooker for four-year-old minds. Harper's, Jan., 1967, pp. 55-61.

39. Rinsland, H. D. A basic vocabulary of elementary school children. New York: Macmillan, 1945.

40. Skinner, B. F. The science of learning and the art of teaching. Harvard educ. Rev., 1954, 24, 86-97.

41. Snow, C. P. The two cultures and the scientific revolution. New York: Cambridge University Press, 1959.

42. Smedslund, J. The acquisition of conservation of substance and weight in children: I. Introduction. Scand. J. Psychol., 1961, 2, 11-20. (a)

43. Smedslund, J. The acquisition of conservation of substance and weight in children: II. External reinforcement of conservation of weight and of the operations of addition and subtraction. Scand. J. Psychol., 1961, 2, 71-84. (b)

44. Smedslund, J. The acquisition of conservation of substance and weight in children: III. Extinction of conservation of weight acquired "normally" and by means of empirical controls on a balance. Scand. J. Psychol., 1961, 2, 85-87. (c)

45. Smedslund, J. The acquisition of conservation of substance and weight in children: IV. Attempt at extinction of the visual components of the weight concept. Scand. J. Psychol., 1961, 2, 153-155. (d)

46. Smedslund, J. The acquisition of conservation of substance and weight in children: V. Practice in conflict situations without external reinforcement. Scand. J. Psychol., 1961, 2, 156-160. (e)

47. Smedslund, J. The acquisition of conservation of substance and weight in children: VI. Practice on continuous vs. discontinuous material in problem situations without external reinforcement. Scand. J. Psychol., 1961, 2, 203-210. (f)

48. Smedslund, J. The acquisition of conservation of substance and weight in children: VII. Conservation of discontinuous quantity and the operations of adding and taking away. Scand. J. Psychol., 1962, 3, 69-77. (g)

49. Torrance, E. P. Guiding creative talent. Englewood Cliffs, N. J.: Prentice-Hall, 1962.

50. Vygotsky, L. S. Thought and language. (Trans. by E. Hanfmann and G. Vakar) Cambridge, Mass.: MIT Press, 1962.

51. Wallach, M. A. Research on children's thinking. In H. W. Stevenson (Ed.), Child psychology. Yearb. nat. Soc. Stud. Educ., 1963, 62 (1), 236-276.

52. Whitehead, A. N. Science and the modern world. New York: Mentor Books, 1948.

53. Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.